

JAPANESE [JP,11-224940,A]

CLAIMS DETAILED DESCRIPTION TECHNICAL FIELD PRIOR ART EFFECT OF THE
INVENTION TECHNICAL PROBLEM MEANS DESCRIPTION OF DRAWINGS DRAWINGS

[Translation done.]

*** NOTICES ***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[The technical field to which invention belongs] this invention has a charge-storage means in the insulator layer between a channel formation field and a control electrode, and relates to the nonvolatile semiconductor memory in which the program of data is possible, and its method of writing in electrically by pouring in or drawing out a charge for the charge-storage means concerned. Specifically, this invention relates to the nonvolatile semiconductor memory and the method of writing in of composition of that time which can forbid writing is lengthened and incorrect writing can be effectively prevented about other storage elements [control electrodes are connected with the storage element concerned, and] to forbid writing, even when the write-in time of a storage element is long. Furthermore, this invention carries out induction of the depletion layer to the semiconductor front face between a storage element and a selection element by the 2nd control electrode between storage elements, and relates to the nonvolatile semiconductor memory which can ensure the ban on writing by controlling the potential of the depletion layer by potential of the 2nd control electrode, and its method of writing in.

[0002]

[Description of the Prior Art] A nonvolatile semiconductor memory element establishes a charge-storage means into the insulator layer made to intervene between the channel formation field (surface fields, such as for example, a semiconductor substrate or a well in which the channel of an element is formed) of a semiconductor, and a control electrode, changes the threshold (generally gate threshold voltage) of a memory device according to the existence or the amount of charges of the charge which exists in this charge-storage means, and is making this threshold correspond to the logic state of a stored data signal conventionally. here, it forms in the inside of the diameter conductor of a granule which was dispersed superficially [single conductive layers such as FG (floating gate), nano crystal, etc.] for a charge-storage means to accumulate a charge, have been arranged, and was insulated, and a nitride, or a nitride and an oxide-film interface -- having -- being spatial (the direction of a flat surface, and the direction of thickness) -- there is a dispersed charge trap [two or more] It responds to the difference in the laminated structure of the gate insulator layer containing this charge trap, and the nonvolatile semiconductor memory element of a kind with various FG types, nano crystal types, MONOS types, MNOS types, etc. is made as an experiment and offered.

[0003] In the nonvolatile semiconductor memory (nonvolatile memory apparatus) from which a majority of these memory devices were arranged and which the memory cell array consisted of, although many kinds of memory cell methods are proposed, there is a NAND type in it as a cell method with the smallest cell size in which large-capacity-izing is possible. A NAND type nonvolatile memory apparatus enables reduction of an efficiency cell area per bit by constituting memory block which carries out the series connection of two or more memory transistors, and is called a NAND train, and sharing one bit contact and a source line between two NAND trains.

[0004] Drawing 13 is the circuit diagram showing the basic composition of the memory cell array of the

conventional NAND type nonvolatile memory apparatus.

[0005] A memory cell array, M11 a-M1na, M11 b-M1nb, M21a, and M21b a sign 100 among drawing 13 A memory transistor, S11a, S12a, S11b, S12b, S21a, S22a, S21b, and S22b A selection transistor, BLaBLb -- in a bit line, WL11-WLn1, and WL21, a bit line selection-signal line, and SG12 and SG22 show a source line selection-signal line, and, as for a word line and SL, BC shows bit contact, as for a source line, and SG11 and SG21 The repeat unit called a string consists of two selection transistors (selector gate) connected to the bit line or the source line, and a NAND train to which the series connection of the n memory transistors (n is the number of 8, 16, and 32 grades) was carried out between both selection transistors. Selection transistor S11a, S11b, S21a, and S21b which were connected to the bit line are controlled by the bit line selection-signal line SG11 or SG21, and selection transistor S12a, S12b, S22a, and S22b which were connected to the source line are controlled by the source line selection-signal line SG12 or SG22. Moreover, memory transistor M11a, M11b and M12a, M12b and M13a, M13b and M1na, and M1nb are controlled by word lines WL11, WL12, WL13, and WLn1, respectively. Similarly, memory transistor M21a and M21b are controlled by the word line WL21.

[0006] Drawing 14 is the direction cross section of a train centering on the NAND train located in the lower left of drawing 13 in the conventional nonvolatile memory apparatus. In addition, it is FG (Floating Gate) about a memory transistor here. The case where it considers as type is illustrated. drawing 14 -- setting -- a sign 2 -- for example, n -- the bit contact by which the semiconductor substrate of type and 4 were punctured by the p type well (p well), and 24 was punctured for a layer insulation layer and 24a by the layer insulation layer 24 -- a hole is shown bit contact -- a hole -- 24a constitutes the aforementioned bit contact BC with the connecting plug embedded at this each -- on the p well 4, the laminating of the tunnel insulator layer 40, the floating gate FG, the insulator layer 42 between the gates, and the control gate CG is carried out, and memory transistor M11 a-M1na is constituted The control gate CG of each memory transistor constitutes a word line WL11 - WLn1, respectively.

[0007] the connection by which the layer which serves as the floating gate FG in a memory transistor with these selection transistor, and the layer used as the control gate CG were prepared in the insulator layer 42 between the gates although the selection transistors SG11, SG12, and SG21 were the same laminated structures as fundamentally as a memory transistor -- it has connected too hastily through the hole As well as the usual monolayer gate, all the gate electrode layers on a gate insulator layer have this potential, and, thereby, the bit line selection-signal lines SG11 and SG21 and the source line selection-signal line SG12 are constituted.

[0008] Thus, source drain impurity-range 6c of a memory transistor and a selection transistor is formed in the surface field of the p well 4 located in the arranged gate inter-electrode space area. alienation of the gate electrode of one selection transistors SG11 and SG21 -- common drain impurity-range 6a is formed in the surface field of the p well 4 located in a space between two strings of the direction of a bit Moreover, source impurity-range 6b which makes the common aforementioned source line SL among other strings who adjoin in the direction of a bit is formed in the surface field of the p well 4 located in the gate electrode outside of the selection transistor SG12 of another side.

[0009] Although drawing 14 shows FG type, even when the nonvolatile memory device to which the charge-storage means was dispersed superficially is used, gate insulation membrane structures differ, and also it is the same as that of drawing 14 fundamentally at the conventional NAND type. Between the p well 4 and a gate electrode (word line), in a MONOS type, the laminating of a tunnel insulator layer, a nitride, and the top oxide film is carried out, and the laminating of a tunnel insulator layer and the nitride is carried out with an MNOS type. Moreover, in the nano crystal type, in the insulator layer on the tunnel insulator layer on the p well 4, the diameter conductor of a granule is dispersed mutually and embedded.

[0010] Explanation [general to an example] of operation is given for the case where a memory transistor makes it correspond to a normally on state and the Nomar Lee OFF state, and binary information is memorized in the NAND type nonvolatile memory apparatus of such composition next. In read-out operation, the voltage clamp of the word line (selection word line) and well to which the cell

(selection cell) to read was connected is carried out to 0V, and the voltage VRG through which all of all selection transistors and the memory transistors connected to word lines other than a selection word line (non-choosing word line) flow is impressed to all selection-signal lines and non-choosing word lines. Only in the potential difference with a well, this voltage VRG is the size by which writing and elimination are not carried out to a memory transistor, for example, is about 5V-7V. If only the bit line (subdevice-bit line) to which the selection cell was connected impresses positive voltage in this state, since all memory transistors other than the cell which reads information are in switch-on, it will be decided whether current will flow [the memory transistor of a selection cell] on a subdevice-bit line by normally on or no MARIOFU, or it will not flow. The existence of this current is detected and the logic state "1" of stored data or "0" is judged.

[0011] Elimination operation is usually performed per block and the high voltage VPP is impressed to all the word lines of a selection block at all the word lines and substrate, or well of 0V and a non-choosing NAND train. Consequently, an electron is drawn out by the substrate from the floating gate, and only the memory transistor of a selection block shifts the threshold voltage of a memory transistor in the negative direction, for example, will be in the elimination state (a logic state corresponds to "1") of normally on.

[0012] On the other hand, program operation of data is performed to the memory transistor package usually connected to the word line to choose in the so-called page unit. Specifically, the middle voltage (path voltage) which is the grade which is in the state which turned off the selection transistor by the side of ON and a source line, and turns on the selection transistor by the side of a bit line although writing is not performed in a non-choosing cell at the high voltage and a non-choosing word line at a selection word line is impressed. At this time, high middle potential is set as the grade by which writing is not carried out to the non-subdevice-bit line by which only the memory transistor which should forbid 0V and a program (for example, "0" data are held) was connected to the subdevice-bit line to which the memory transistor which should be programmed (for example, "0" data are memorized) was connected by the potential difference with the potential by the above-mentioned impression high voltage. Consequently, the threshold voltage of a selection memory transistor shifts only the selection memory transistor which should be programmed in the right direction by pouring in an electron into the floating gate, and it is high, for example, will be in the write-in state of no MARIOFU from the above-mentioned elimination state.

[0013] In this write-in operation, since the operating time is ****(ed) by the charge and discharge of the bit line which usually has high load-carrying capacity, the technology which separates only a non-choosing NAND train from a bit line by the selector gate is indicated by the patent public presentation official report common No. 97455 [six to] by setting it as the main purposes to reduce the voltage which drives a bit line and to mitigate the burden of a pressure-up system circuit. With this technology, it is supposed that the applied voltage of a non-subdevice-bit line is reduced to the grade which lengthened the threshold from the applied voltage of a selector gate, and a non-choosing NAND train will be separated from a bit line. Consequently, since the channel potential (write-in retarding voltage) of a non-choosing NAND train is set up by the automatic pressure up according [henceforth] to path voltage or program voltage, the technology indicated by the above-mentioned reference is widely known as self boost technology now.

[0014]

[Problem(s) to be Solved by the Invention] However, when this self boost technology is applied to the conventional nonvolatile memory apparatus which showed basic composition with the cross section of drawing 14 , there are some problems shown below.

[0015] Although the tunnel insulator layer is further thin-film-ized in non-volatile memory as reduction of supply voltage and cell detailed-ization progress It is pointed out that the thickness limitation which especially originated in stress leak with FG type exists. As for the thickness limitation, drawing speed may become slow if a low-battery drive progresses while 6nm is actually set to about 8nm (refer to the Nikkei micro device January issue and the February issue, and 1997) and the scaling of the tunnel thickness has not been carried out by it with a theoretical value, either. The 1st problem is that the

memory transistor of the non-choosing cell connected to the same word line as the selection cell which is writing in data by such reason when drawing speed is slow is [incorrect-] easy to be written in.

According to the reference (open official report) mentioned previously, it is shown by the graph that a threshold rises, if time to change a non-choosing NAND train into a write-protected state effectively, and no threshold shift be has a certain fixed upper limit, for example, 10msec(s) are exceeded.

Therefore, if write-in time becomes long exceeding this upper limit, by the culmination of writing, a charge will be injected into the memory transistor of the non-choosing cell connected to the same selection word line, and it will be written in somewhat.

[0016] Although it is fundamentally the same also in a MONOS type, an MNOS type, etc., since [this 1st problem] the charge trap is dispersed, and it compares with FG type and excels in a MONOS type and an MNOS type at the scaling nature of a gate insulator layer, it is performing scalings, such as a gate insulator layer, with the fall of program voltage, and drawing speed cannot fall easily with low-battery-izing at the time of writing, and it is more advantageous than FG type at this point.

[0017] It is mentioned that the shift amount (V_{th} window width of face) of the threshold of the write-in state which is in a reduction-ized inclination with detailed-izing of an element as the 2nd technical problem, and an elimination state does not suit low-battery-ization of program voltage. That is, voltage which makes the non-choosing memory transistor of a NAND train turn on especially at the time of read-out if V_{th} window width of face is not reduction-ized to some extent in the NAND type non-volatile memory by which the tunnel insulator layer of a storage element etc. is optimized, either cannot be made low so that it may be conventionally programmed on low voltage, but it becomes easy to produce the non-choosing memory transistor weakly written in according to the variation in a threshold. Also at this point, generating plain-gauze ***** has incorrect writing.

[0018] this invention is made in view of such the actual condition, and aims at offering the nonvolatile semiconductor memory of structure which prevents effectively the incorrect writing to the non-choosing cell in accordance with reduction of program voltage. Moreover, this invention can be suitably carried out in this nonvolatile semiconductor memory, and sets it as other purposes to offer how to write in the nonvolatile semiconductor memory which can prevent effectively the incorrect writing to the non-choosing cell in accordance with reduction of program voltage.

[0019]

[Means for Solving the Problem] In order to solve the trouble of the conventional technology mentioned above and to attain the above-mentioned purpose, in the nonvolatile semiconductor memory of this invention For example, the NAND train of composition etc. is conventionally set in cell composition to which the series connection of the transistor is carried out. It considers as composition without an impurity addition field (source drain field) common between the transistor. We decided that it is made to depletion-ize by capacity coupling, this writes in by the non-choosing memory transistor side by which the semiconductor region concerned was especially connected to the same word line as a selection memory transistor, and induction of the retarding voltage is carried out. Namely, the nonvolatile semiconductor memory of this invention Voltage is impressed to the 1st control electrode by which the laminating was carried out through the insulator layer including a charge-storage means on the channel formation field of a semiconductor. The storage element which memorizes information by pouring in a charge electrically to the aforementioned charge-storage means, or drawing out a charge from the charge-storage means concerned, It is the nonvolatile semiconductor memory which has the element which transmits predetermined potential to the aforementioned channel formation field in pouring of the aforementioned charge, or the case of drawing. Capacity coupling is carried out to the semiconductor region between the elements which transmit the aforementioned storage element and the aforementioned predetermined potential. The 2nd control electrode which performs formation of the channel between elements or a depletion layer and control of potential to the semiconductor region concerned is prepared through the insulator layer on the semiconductor region between the elements which transmit a storage element and predetermined potential concerned. The elements which transmit this predetermined potential may be other storage elements, such as a NAND type, and may be selection elements which control connection between common wiring and the aforementioned storage element, and interception

between the storage elements of the plurality of a line writing direction or the direction of a train.

[0020] Preferably, the 2nd control electrode of the above is prepared in common among two or more aforementioned storage element trains which have a wrap plate configuration through an insulator layer to the 1st control electrode of the above, and the aforementioned charge-storage means, and are located in a line with a line writing direction at least. Moreover, two or more selection elements which control the connection between common wiring and the aforementioned storage element, and interception, respectively between the desirable storage elements of the plurality of a line writing direction or the direction of a train, The applied voltage of the control electrode of two or more selection elements concerned, the 1st control electrode of two or more aforementioned storage elements, the 2nd control electrode, and the aforementioned wiring is controlled, respectively. While forming an inversion layer in the aforementioned semiconductor region between elements which adjoined the selective-memory element which writes in and supplying the aforementioned predetermined potential through the selection element of the switch-on corresponding to the inversion layer concerned The 1st control electrode intercepts the selection element which corresponds the semiconductor region between elements which adjoined the storage element [**** / un-] by which common connection was made with the aforementioned selective-memory element, depletion-izes as a suspension state electrically, and has further the bias control means which carry out induction of the write-in predetermined retarding voltage to the semiconductor region between the elements concerned.

[0021] How to write in the nonvolatile semiconductor memory of this invention On the channel formation field of a semiconductor, arrange two or more storage elements to which it comes to carry out the laminating of the control electrode through an insulator layer including a charge-storage means in the shape of a matrix, and a memory cell array is constituted. It is how to write in the nonvolatile semiconductor memory which pours a charge into the aforementioned charge-storage means electrically, or draws out a charge from the charge-storage means concerned, and memorizes information by impressing program voltage to the aforementioned control electrode. The inversion layer by the minority carrier is formed in the semiconductor region which adjoined the channel formation field about the aforementioned storage element which writes in. It writes in by holding the inversion layer concerned with the predetermined potential to which writing is performed by the potential difference with the potential by the above-mentioned program voltage impression. About other storage elements which do not write in the information for which control electrodes were connected with the aforementioned storage element, a depletion layer is formed during the above-mentioned writing at the semiconductor region which adjoined the channel formation field, and induction of the write-in retarding voltage is carried out.

[0022] By such nonvolatile semiconductor memory and its method of writing in, of capacity coupling of the above 1st with which predetermined potential was impressed, respectively, and the 2nd control electrode, a minority carrier can draw near to a channel formation field and the semiconductor region between elements, and an inversion layer is formed in a memory transistor for every singular number or included transistor train. However, when there are two or more transistor trains, a part for the surface flank of the semiconductor region between the element is electrically connected to wiring, and in case a channel is formed in an adjoining field, time difference arises in the case where a certain potential is supplied, and the case where it is separated from the wiring which supplies potential. In such control, if it explains by the NAND [which forbids writing, for example using self boost technology] type case, since the selection transistor connected to NAND **** of a selection block turns on and bit line potential (for example, 0V) is supplied, a minority carrier is quickly supplied from a bit line, and, as for the NAND train of this selection block, a channel is formed early. On the other hand, since the selection transistor by the side of a bit line intercepts in a non-choosing block, although fixed for example, to grounding potential, since the voltage clamp of the surface field of a channel formation field edge is not carried out, the well which has a channel formation field on a front face does not have the rapid source of supply of a minority carrier. For this reason, as compared with the case where predetermined potential (for example, grounding potential) is supplied from the bit line, formation of a channel is slow and the state where it depletion-ized is held for a long time. At this time, the semiconductor region between

elements of a non-choosing block becomes possible [pull-up of the potential being carried out by capacity coupling with the 2nd control electrode, consequently carrying out induction of the write-in predetermined retarding voltage to the semiconductor region between the elements concerned, as a result an adjoining channel formation field].

[0023] even if a non-choosing block is not connected to wiring of a bit line etc. in the former -- a NAND train -- on the way -- since it is alike and the impurity diffusion layer (source drain field) of a channel formation field and a reverse conductivity type is prepared by turns across the channel formation field at intervals of predetermined, although this functions as a source of supply of a minority carrier and is later than channel formation of a selection block, it will still be comparatively alike and a channel will be formed in a quick stage. Moreover, it was difficult for the potential difference between the elements which continued by the channel to decrease, to set up the absolute value greatly from the potential of the channel formation field of the storage element which adjoins the potential of the channel formation field of a storage element to carry out the ban on writing, and to ensure the ban on writing.

[0024] On the other hand, in the nonvolatile semiconductor memory of this invention, there is no impurity diffusion layer which acts as such a carrier source of supply, and formation of the channel (inversion layer) in the channel formation field which adjoins the portion can be made later than before, consequently write-protected time becomes long. That is, the time margin to incorrect writing is expanded. Moreover, that this channel formation can be made late means that a voltage margin with the threshold value which incorrect writing produces is expanded about the potential difference concerning a tunnel insulator layer, if it sees at a certain time. Furthermore, since induction of the depletion layer is carried out to the semiconductor-region front face between a storage element and a selection element between storage elements and the surface potential can be controlled by potential of the 2nd control electrode, it becomes possible to set up greatly the absolute value of the surface potential of the channel formation field of a storage element to carry out the ban on writing, and the ban on writing can be ensured.

[0025] When the charge-storage means is especially dispersed within the flat surface, the above-mentioned voltage margin can be enlarged further. In FG type, if applied voltage of the 2nd control electrode is enlarged, although the potential of a depletion layer (semiconductor region between elements) will rise, charge-storage means (floating gate) potential will also rise to some extent by capacity coupling. On the other hand, when the charge-storage means is dispersed within the flat surface, since capacity coupling of this and the 2nd control electrode is hardly carried out, they can reduce quickly the potential difference concerning the tunnel insulator layer which causes incorrect writing.

[0026]

[Embodiments of the Invention] this invention considers a storage element pouring in a charge electrically to charge-storage meanses (for example, the floating gate, a charge trap, etc.) using capacity coupling of the control electrode, or drawing out as basic operation, and the storage element is widely applied to the nonvolatile semiconductor memory which has the cell composition by which a series connection is carried out to other storage elements or a selection transistor. if it is such a nonvolatile semiconductor memory -- the kind of storage element -- limitation -- there is nothing -- the usual FG (Floating Gate) MNOS (Metal-Nitride-Oxide Semiconductor) besides type, and MONOS (Metal-Oxide-Nitride-Oxide Semiconductor) -- it is further applicable to a nano crystal type etc. Moreover, there is no limitation in a memory cell method, and it can apply to a NOR type (the thing by which the bit line and the source line were hierarchized, for example, a DINOR type, and an AND type are included) besides a NAND type. Hereafter, the operation gestalt of the nonvolatile semiconductor memory concerning this invention and its manufacture method is explained in detail based on a drawing by making into an example the NAND type nonvolatile memory apparatus which has MONOS, an MNOS type, and FG type FET as a storage element.

[0027] The 1st operation gestalt book operation gestalt is related with a MONOS type. Drawing 1 is the circuitry view of the memory cell array of the NAND type nonvolatile memory apparatus concerning this operation gestalt. Moreover, drawing 2 is cross-section structural drawing of the NAND train (it

corresponds to the NAND train at the lower left of drawing 1) concerning this operation gestalt.

[0028] In drawing 1 , two transistor trains (string) are connected to each bit lines BLa and BLb, respectively. The 1st string connected to the bit line BLa consists of selection transistor S11a connected to the bit line BLa, selection transistor S12a connected to the source line SL, and n memory transistor M11 a-M1na(s) by which the series connection was carried out between both selection transistors. The 2nd string connected to the bit line BLa consists of selection transistor S21a connected to the bit line BLa, selection transistor S22a connected to the source line SL, and n memory transistor M21a-- by which the series connection was carried out between both selection transistors. Similarly, two strings are connected also to the bit line BLb, and the 1st string consists of selection transistor S11b connected to the bit line BLb, selection transistor S12b connected to the source line SL, and n memory transistor M11 b-M1nb(s) by which the series connection was carried out between both selection transistors. The 2nd string connected to the bit line BLb consists of selection transistor S21b connected to the bit line BLb, selection transistor S22b connected to the source line SL, and n memory transistor M21b-- by which the series connection was carried out between both selection transistors.

[0029] Both selection transistor S11a and S11b are controlled by the bit line selection-signal line SG11, and both selection transistor S12a and S12b are controlled by the source line selection-signal line SG12. Similarly, both selection transistor S21a and S21b are controlled by the bit line selection-signal line SG21, and both selection transistor S22a and S22b are controlled by the source line selection-signal line SG22. Moreover, memory transistor M11a, M11b and M12a, M12b and M13a, M13b, -- and M1na, and M1nb are controlled by word lines WL11, WL12, and WL13, --, WL1n, respectively. Similarly, memory transistor M21a and M21b are controlled by the word line WL21.

[0030] Thus, a p type well (p well 4) is formed in the front-face side in the n semiconductor substrate 2 of type, and the transistor train is making it arrange to the front-face side of the p well concerned in each string by whom interconnection is done, as shown in the cross section of drawing 2 .

[0031] memory transistor M11 a-M1na -- the p well 4 top -- the tunnel insulator layer 8, a nitride 10, and the top oxide film 12 -- a shell -- the gate insulator layer is constituted Moreover, the laminating of the word line WL11 - WL1n which consists of a polysilicon contest layer 16 and a refractory-metal silicide layer 18 on it on a gate insulator layer is carried out. This word line corresponds to the "1st control electrode" of this invention. Moreover, p well surface portion between a "channel formation field", and the 1st control electrode concerned corresponds to "the semiconductor region between elements" in this invention. [in / this invention / in p well surface portion of the 1st control-electrode lower part which consists of a polysilicon contest layer and a refractory-metal silicide layer]

[0032] Selection transistor S11a, S12a, etc. consist of usual MOSFETs. Therefore, the gate insulator layer 14 consists of only silicon oxides. Each gate electrode layer of selection transistor S11a and S12a constitutes the bit line selection-signal line SG11 and the source line selection-signal line SG12, respectively.

[0033] One of the structural features of the non-volatile memory of this operation form is that the impurity range (source drain field 6c in drawing 14 which shows the conventional example) of this p well 4 and a reverse conductivity type is not formed in the surface field (semiconductor region between elements) of the p well 4 between a memory transistor and a selection transistor between memory transistors. Other strings who adjoin the one side of the impurity range of a bit of other reverse conductivity types, i.e., the direction, common drain impurity-range 6a of a reverse conductivity type, and source impurity-range 6b (source line SL) of a reverse conductivity type common among the contiguity strings of another side are prepared as usual.

[0034] Other structural features of the non-volatile memory of this operation form are that the 2nd control electrode 22 is formed through the layer insulation layer 20 which consists of a silicon oxide on a transistor train. It is formed in the shape of one plate for every transistor train, applying [this / 2nd / 22] it above the selection transistor of another side from the upper part of the selection transistor of one of these. And the 2nd control electrode 22 is close through the layer insulation layer 20 to the semiconductor region between elements between each memory transistor or between a memory transistor and a selection transistor, and controls potential in the channel between elements, or the

formation row of a depletion layer in the p well 4 surface portion (semiconductor region between elements) of the lower part according to the applied voltage of the control electrode 22 concerned.

[0035] the contact by which it was comparatively alike, the thick layer insulation layer 24 was formed all over including the 2nd control-electrode 22 top, and the bit contact BC was punctured on the aforementioned drain impurity-range 6a about this layer insulation layer 24 -- a hole -- it is formed by making adhesion layers, such as Ti/TiN, intervene and embedding the inside of 24a with metal plugs, such as W In the form connected with a transistor train by the bit contact BC, the bit line BL_a is wired on the layer insulation layer 24. Although especially the bit line BL_a does not illustrate, it usually has the three-tiered structure which sandwiched the upper and lower sides of the main wiring layers, such as aluminum, with an acid-resisting layer (or protective layer) and barrier metal.

[0036] In such a NAND type nonvolatile memory apparatus of composition, since there is no impurity range among other transistors which adjoin a memory transistor, there is an advantage on which incorrect writing is effectively prevented so that it may mention later.

[0037] Below, the manufacture method is explained. Drawing 3 - drawing 6 are the cross sections showing each manufacture process of the nonvolatile memory apparatus of composition of having mentioned above.

[0038] In advance of this drawing 3, the isolation field by LOCOS or the trench is first formed in the surface portion of the prepared substrate (n-type-semiconductor substrate 2). Although this isolation field does not appear in a cross section, it is formed in the direction of a bit in the shape of [long] an parallel stripe each transistor train and by turns, and, thereby, the isolation of a line writing direction is mainly attained.

[0039] Subsequently, as shown in drawing 3 (a), the p well 4 is formed in the surface portion in a substrate with ion-implantation. In addition, the SOI layer formed through the insulating layer on the epitaxial growth phase and the substrate (possible also except semiconductors, such as a quartz and silicon on sapphire) can be substituted for this p well 4. The tunnel insulator layer 8 is formed by oxidizing thermally p well 4 front face, for example. The tunnel insulator layer 8 is good also considering the surface portion or all thermal oxidation silicon as a nitriding silicon oxide by carrying out heat nitriding treatment of the thermal oxidation silicon besides a silicon oxide. On the tunnel insulator layer 8, the nitride 10 which consists of a silicon nitride is formed by CVD etc.

[0040] Next, as shown in drawing 4 (b), the resist pattern RP which carries out a coat only to the portion in which a memory transistor train is formed is formed, the nitride 10 of a ground is *****ed on a mask in this, and the nitride of a portion in which a selection transistor is formed is removed. It is desirable to leave the tunnel insulator layer 8 in part at least, and for an etching damage to be made not to be introduced into a substrate side in the case of this etching.

[0041] It oxidizes thermally in continuing drawing 4 (c). When it leaves the tunnel insulator layer 8 in part at least at a previous process, it is desirable to remove completely by wet processing of pretreatment of thermal oxidation etc. While the thermal oxidation film which becomes the selection transistor section which the well 4 expressed mostly with the gate insulator layer 14 of a selection transistor by this thermal oxidation is formed thickly, on a nitride 10, the thermal oxidation film used as the top oxide film 12 of a memory transistor is formed thinly. At this time, a nitride 10 carries out film decrease and the nitriding thickness of a final request is obtained.

[0042] In drawing 5 (d), the laminating of the polysilicon contest layer 16 by which the impurity was introduced and electric-conduction-ized on the thermal oxidation film 12 and 14, and the refractory-metal silicide layer 18 is carried out by the usual polycide forming method.

[0043] In continuing drawing 5 (e), using the resist pattern RP, the cascade screen of a ground is *****ed on a mask in this, and simultaneous formation of the parallel stripe-like word line WL11 - WL1n, etc. and the selection-signal line SG11 and SG12 grade is carried out. Although dirty-off is carried out for all in this drawing until a silicon layer expresses, in order to suppress the damage introduction by the side of a substrate, it may leave the oxide film of the lowest layer partially to all fields, and wet etching may remove later.

[0044] As shown in drawing 6 (f) after removing the resist pattern RP to the next, the 1st layer

insulation layer 20, such as a silicon oxide, is thinly formed on the whole surface. This membrane formation is formed by which method of thermal oxidation or CVD. When a word line etc. becomes with contest polysilicon, the oxidizing [thermally] method can be adopted. When a word line etc. is a polycide, it is based on CVD. Then, the conductive layer used as the 2nd control electrode 22 is formed so that it may moreover embed gate inter-electrode on the whole surface, and etching processing is carried out, using as a mask the resist pattern which does not illustrate this. The 2nd control electrode 22 formed by this needs to cover inter-electrode [of a memory transistor] at least, in order to control potential in the channel between elements, or the formation row of a depletion layer. The composition which embeds, forms in the shape of a line, and is connected electrically in the part which was separated for every transistor inter-electrode space, and which is not illustrated is sufficient as the 2nd control electrode. With the resist pattern of this 2nd control electrode 22 or its formation mask attached, an ion implantation is performed for this on a mask, and simultaneous formation of the source impurity-range 6b which becomes drain impurity-range 6a and the source line SL at the surface portion of the p well 4 is carried out by using the 1st layer insulation layer 20 as a through film.

[0045] after that, it is shown in drawing 2 -- as -- the 2nd thick layer insulation layer 24 -- depositing -- flattening -- carrying out -- bit contact -- a hole -- the nonvolatile memory apparatus concerned is completed by performing puncturing of 24a, embedding of a connecting plug, wiring of a bit line, etc.

[0046] Next, in the NAND type nonvolatile memory apparatus of such composition, data writing of a memory transistor, elimination, and explanation of read-out of operation are given. Although this invention can be applied when this memory transistor memorizes the multiple-value information other than binary information, explanation of operation between two states of normally on and no MARIOFU is given here. In addition, fundamental operation is the same from carrying out by shifting gradually the word line voltage or the bit line at the time of a program or read-out in the right direction in the case of a multiple value.

[0047] First, in read-out operation, the voltage clamp of the word line (selection word line) and well to which the cell (selection cell) to read was connected is carried out to 0V, and the voltage VRG through which all of all selection transistors and the memory transistors connected to word lines other than a selection word line (non-choosing word line) flow is impressed to all selection-signal lines and non-choosing word lines. This voltage VRG is a size by which writing and elimination are not carried out to a memory transistor only in the potential difference with a well. Moreover, in the NAND train containing the cell which reads the voltage impressed to the 2nd control electrode of the above, the voltage in which the channel between the element is formed is chosen. If only the bit line (subdevice-bit line) to which the selection cell was connected impresses positive voltage in this state, since all memory transistors other than the cell which reads information are in switch-on, it will be decided whether current will flow [the memory transistor of a selection cell] on a subdevice-bit line by normally on or no MARIOFU, or it will not flow. The existence of this current is detected and the logic state "1" of stored data or "0" is judged.

[0048] As usual, it is carried out by carrying out a memory cell array package, and elimination operation is in a block unit or the state which turned off all the selection transistors, and impresses the high voltage VPP to all the word lines of an elimination unit at all the word lines and substrate, or well of 0V and a non-choosing NAND train. Consequently, only the memory transistor of a selection block is poured into the trapping level (charge trap) to which an electron hole exists near the interface of a nitride and a nitride, and an oxide film from a substrate side, and shifts the threshold voltage of a memory transistor in the negative direction, for example, will be in the elimination state (a logic state corresponds to "1") of normally on.

[0049] On the other hand, it is performed to the memory transistor package connected to the word line by program operation of data in the so-called page unit. Although writing is concrete usually performed from the page near the source side of a selection block line, first, it is in the state which turned off the selection transistor by the side of a source line, and for example, the supply voltage VDD grade (for example, about 5V) is impressed to the subdevice-bit line at the gate (bit line selection-signal line) and the non-subdevice-bit line of a selection transistor by the side of 0V and a bit line. And the high program

voltage VPP is impressed to the voltage and the selection word line which are the grade through which a non-choosing cell flows at a non-choosing word line. Then, since the selection transistor connected to NAND **** of a selection block turns on and bit line potential (for example, 0V) is supplied, a minority carrier is quickly supplied from a bit line, and, as for the NAND train of this selection block, a channel is formed early. Consequently, an electron is poured into the trapping level to which only the selection memory transistor which should be programmed exists the inside of a nitride, and near the interface of a nitride and an oxide film, and the threshold voltage of a selection memory transistor is shifted in the right direction, and will be high for example, be in the write-in state of no MARIOFU from the above-mentioned elimination state.

[0050] On the other hand, since it changes to a cut off state from switch-on with a non-choosing block by the slight potential rise whose selection transistor by the side of the bit line lengthened the threshold voltage of a selector gate from the applied voltage of a bit line selection-signal line Although fixed to grounding potential, from the voltage clamp of the surface field of a channel formation field edge not being carried out, the well itself which has a channel formation field on a front face does not have a source drain impurity range common between elements like before, and it does not have the rapid source of supply of a minority carrier. For this reason, as compared with the selection NAND train side to which predetermined potential (for example, grounding potential) is supplied from the bit line, formation of a channel is slow and the state where it depletion-ized is held for a long time (fixed time). And while this depletion layer is formed, pull-up of the potential is carried out by capacity coupling with the 2nd control electrode 22, consequently induction of the write-in predetermined retarding voltage is carried out to the semiconductor region between the elements concerned, as a result an adjoining channel formation field, and, as for the semiconductor region between elements of a non-choosing NAND train, incorrect write-in prevention is attained.

[0051] Thus, with this operation form, since the self boost of the semiconductor region between elements in which the source drain impurity range was formed by the former is carried out in the state where it depletion-ized by capacity coupling with the 2nd control electrode 22, a time margin for the impression pulse (write-in pulse) of the program voltage VPP to end to during this period arises, namely, write-protected time becomes long. The time margin to incorrect writing is expanded as a result. Moreover, that this channel formation can be made late means that a voltage margin with the threshold value which incorrect writing produces is expanded about the potential difference concerning a tunnel insulator layer, if it sees at a certain time. Especially, in this example, since the charge-storage means is dispersed within the flat surface, the above-mentioned voltage margin can be enlarged further. Namely, if applied voltage of the 2nd control electrode is enlarged, although the potential of a depletion layer (semiconductor region between elements) rises and charge-storage means (floating gate) potential will also rise to some extent by capacity coupling with the 2nd control electrode in FG type In this example, since the charge-storage means is dispersed within the flat surface, since capacity coupling of this and the 2nd control electrode is hardly carried out, they can reduce quickly the potential difference concerning the tunnel insulator layer which causes incorrect writing.

[0052] In a nonvolatile memory apparatus, while program operation repeats a kind of read-out operation which generally verifies whether the desired threshold was obtained by the program by the write-in pulse, it performs. That is, if a desired threshold is obtained in the verification after a program, about the memory transistor, the following program will be performed setting up a write-protected state by self boost operation, and a threshold will be verified again. About the thresholds of all the memory transistors by which this was connected to the word line, it carries out repeatedly until a desired value is acquired.

[0053] By the nonvolatile memory apparatus of this operation form, and its method of writing in, the control which wrote in by the ability lengthening write-protected time, and made pulse width larger than before is possible, and there is an advantage to which the flexibility of control of write-in pulse width increases in the multiple-value memory technology in which the permission width of face of the threshold finally controlled especially is narrow. Moreover, even when it writes in from the reason of the scaling of an element not suiting low-battery-ization when program voltage is low-battery-ized and time

becomes late, incorrect writing can be prevented effectively. Furthermore, when program voltage is in a low-battery-ized inclination and reduction-ization of V_{th} window width of face does not fulfill this, in order to make all the non-choosing memory transistors at the time of read-out turn on. Although the fall of the path voltage impressed to the gate cannot be advanced easily, that channel formation cannot be easily carried out by this invention. In the former, the meaning can also prevent incorrect writing effectively with the operation in which incorrect writing with the high path voltage which will be in a weak write-in state cannot occur easily.

[0054] The 2nd operation gestalt book operation gestalt is about an MNOS type nonvolatile memory apparatus. The circuitry of a fundamental cell array shown in drawing 1 and operation do not change with the 1st operation gestalt except the nonvolatile memory apparatus of this operation gestalt having changed the memory transistor into the MNOS type from the MONOS type of the 1st operation gestalt. Therefore, only the cross-section structure and the manufacture method of a NAND train are explained, and others omit here. Moreover, the composition which overlaps in cross-section composition attaches the same sign, and detailed explanation is not given.

[0055] Drawing 7 is the direction cross section of a bit of the NAND train in the nonvolatile memory apparatus concerning the 2nd operation gestalt of this invention. The gate insulator layer comes to carry out [na / memory transistor M11 a-M1/ of this operation gestalt] the laminating of the tunnel insulator layer 30 and the nitride 32 on the p well 4. Moreover, like the 1st operation gestalt, although the gate insulator layer 34 of a selection transistor may be made to constitute only from a film of a silicon oxide, it is taken as the cascade screen of a silicon-oxide film and a silicon nitride film here. In this case, it is comparatively alike, and since [whose lower layer silicon oxidation thickness is about 10nm] it is thick, it does not function as a memory device. Since this NAND type nonvolatile memory apparatus does not have an impurity range like the 1st operation gestalt among other transistors which adjoin a memory transistor, it has the advantage on which incorrect writing is prevented effectively.

[0056] Drawing 8 - drawing 11 are the cross sections showing each manufacture process of the nonvolatile memory apparatus of composition of having mentioned above. First, the about 10nm gate oxide film 5 thick in comparison is formed by forming an isolation field and the p well 4 in the prepared substrate 2 like the 1st operation gestalt, and oxidizing thermally p well 4 front face, for example (drawing 8 (a)).

[0057] As shown in drawing 9 (b), the resist pattern RP is formed on this gate oxide film 5, and etching which made this the mask removes the gate oxide film 5 of the memory transistor section. In continuing drawing 9 (c), the tunnel insulator layer 30 of a memory transistor is formed by the oxidizing [thermally] method. At this time, as for the gate oxide film 5, oxidation thickness increases slightly. In addition, once the tunnel insulator layer 30 forms a thermal oxidation silicon film, it may carry out the heat nitriding of the surface portion at least. Moreover, a nitride 32 is thickly deposited in comparison with dozens of nm by CVD on the tunnel insulator layer 30.

[0058] Henceforth, the polycide which consists of a polysilicon contest layer 16 and a refractory-metal silicide layer 18 is henceforth formed like the 1st operation gestalt (drawing 10 (d)), it *****s the cascade screen of a ground on a mask in this using the resist pattern RP, and simultaneous formation of a word line WL11 - WL1n, etc. the selection-signal line SG11, and the SG12 grade is carried out (drawing 10 (e)). Moreover, after forming the 1st layer insulation layer 20 thinly after removing the resist pattern RP (drawing 11 (f)) and forming the 2nd control electrode 22 on this (drawing 11 (g)), the non-volatile memory concerned is completed through many processes, such as deposition of the 2nd layer insulation layer 24, formation of the bit contact BC, and wiring of a bit line BLA.

[0059] Also in the nonvolatile memory apparatus of this operation gestalt, it has the same advantage as the 1st operation gestalt. That is, like before, since it does not have the impurity range of this and a reverse conductivity type, in the ban on the writing of a non-choosing NAND train, the write-protected time can be lengthened at p well internal-surface side within a NAND train with a self boost. Therefore, it is, when drawing speed is slow and reduction-ization of V_{th} window width of face does not fulfill demand with this low-battery-izing by low-battery-ization of program voltage, and incorrect writing can prevent effectively. Moreover, the flexibility of control of write-in pulse width increases.

[0060] The 3rd operation gestalt book operation gestalt is about FG type nonvolatile memory apparatus. The circuitry of a fundamental cell array shown in drawing 1 and operation do not change with the 1st operation gestalt except the nonvolatile memory apparatus of this operation gestalt having changed the memory transistor into FG from the MONOS type of the 1st operation gestalt.

[0061] Drawing 12 is the direction cross section of a bit of the NAND train in FG type nonvolatile memory apparatus concerning a **** 3 operation gestalt. As shown in this drawing 12, the floating gate 42 which replaces with the nitride of the 1st operation gestalt, for example, consists of contest polysilicon between the tunnel insulator layer 8 and the middle insulator layer 40 (equivalent to the top oxide film of the 1st operation gestalt) intervenes. The middle insulator layer 40 consists of a silicon oxide or an ONO film like the top oxide film of the 1st operation gestalt. Other composition which attached the same sign is the same as the 1st operation gestalt.

[0062] Such a manufacture method of FG type nonvolatile memory apparatus of composition Not the nitride 10 but a polysilicon contest film is formed by drawing 3 in the 1st operation gestalt, and patterning of this is carried out by drawing 4 (b), It can carry out like drawing 3 which shows the manufacture method of the 1st operation gestalt - drawing 6 except forming a middle insulator layer like the top oxide film 12 by drawing 4 (c), and carrying out patterning of the middle insulator layer concerned and the polysilicon contest film simultaneously at the time of electrode processing of drawing 5 (e).

[0063] Also in the nonvolatile memory apparatus of this operation gestalt, it has the same advantage as the 1st operation gestalt. That is, since it does not have an impurity diffusion field between elements like before, it sets write-protected, write-protected time can be lengthened, and the incorrect writing of a non-choosing NAND train can prevent effectively, and it has the advantage of many by the self boost which said that the flexibility of control of write-in pulse width increased.

[0064]

[Effect of the Invention] According to the nonvolatile semiconductor memory concerning this invention, and its method of writing in, the incorrect writing in accordance with the fall of program voltage is prevented effectively, and the flexibility at the time of saying that late write-in control, for example, write-in pulse width, is lengthened increases. Therefore, it becomes possible the nonvolatile semiconductor memory in which reliability and the property were excellent even when detailed-izing and low-battery writing-ization was promoted, and to offer the method of writing in.

[Translation done.]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-224940

(43) 公開日 平成11年(1999) 8月17日

(51) Int.Cl.⁸

識別記号

F I

H 0 1 L 27/115

H 0 1 L 27/10

4 3 4

G 1 1 C 16/04

G 1 1 C 17/00

6 2 2 Z

H 0 1 L 21/8247

H 0 1 L 29/78

3 7 1

29/788

29/792

審査請求 未請求 請求項の数14 O L (全 16 頁)

(21) 出願番号 特願平10-121461

(71) 出願人 000002185

(22) 出願日 平成10年(1998) 4月30日

ソニー株式会社

東京都品川区北品川6丁目7番35号

(31) 優先権主張番号 特願平9-336202

(72) 発明者 青笹 浩

東京都品川区北品川6丁目7番35号 ソニー株式会社内

(32) 優先日 平9(1997)12月5日

(72) 発明者 林 豊

東京都品川区北品川6丁目7番35号 ソニー株式会社内

(33) 優先権主張国 日本 (J P)

(72) 発明者 藤原 一郎

東京都品川区北品川6丁目7番35号 ソニー株式会社内

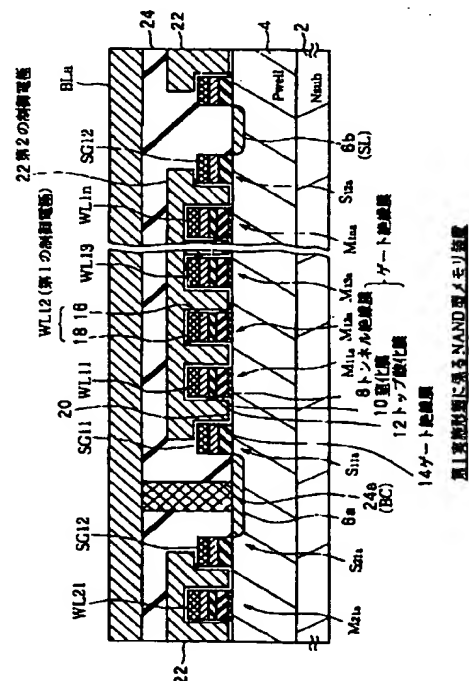
(74) 代理人 弁理士 佐藤 隆久

(54) 【発明の名称】 不揮発性半導体記憶装置及びその書き込み方法

(57) 【要約】

【課題】 プログラム電圧の低減に伴う非選択セルへの誤書き込みを有効に防止する。

【解決手段】 半導体のチャネル形成領域上に電荷蓄積手段を含む絶縁膜を介して積層された第1の制御電極（例えば、ワード線WL11～WL1n）に電圧を印加し、電荷蓄積手段への電荷の注入又は引き抜きを情報記憶の基本動作とする記憶素子M11a～M1naと、その電荷の注入又は引き抜きの際に記憶素子のチャネル形成領域に所定の電位を伝達する素子（例えば、他の記憶素子又は選択素子）とを有する。また、記憶素子と所定の電位を伝達する素子との間の半導体領域と容量結合し、当該半導体領域に対し素子間チャネルまたは空乏層の形成と電位の制御を行う第2の制御電極22が、当該素子間不純物領域上に絶縁膜20を介して設けられている。



1

【特許請求の範囲】

【請求項 1】半導体のチャネル形成領域上に電荷蓄積手段を含む絶縁膜を介して積層された第 1 の制御電極に電圧を印加し、前記電荷蓄積手段に対し電荷を電気的に注入し又は当該電荷蓄積手段から電荷を引き抜くことにより情報を記憶する記憶素子と、前記電荷の注入又は引き抜きの際に前記チャネル形成領域に所定の電位を伝達する素子とを有する不揮発性半導体記憶装置であって、前記記憶素子と前記所定電位を伝達する素子との間の半導体領域と容量結合し、当該半導体領域に対し素子間チャネルまたは空乏層の形成と電位の制御を行う第 2 の制御電極が、当該記憶素子と所定電圧を伝達する素子との間の半導体領域上に絶縁膜を介して設けられている不揮発性半導体記憶装置。

【請求項 2】前記チャネル形成領域と、前記記憶素子と前記所定電位を伝達する素子との間の半導体領域とは、同一の不純物添加領域から構成されている請求項 1 に記載の不揮発性半導体記憶装置。

【請求項 3】前記記憶素子が複数、行列状に配置されてメモリセルアレイが構成され、前記所定電位を伝達する素子は、行方向又は列方向の複数の記憶素子間で共通な配線と前記記憶素子との接続および遮断を制御する選択素子である請求項 1 に記載の不揮発性半導体記憶装置。

【請求項 4】前記記憶素子が複数、行列状に配置されてメモリセルアレイが構成され、前記所定電位を伝達する素子は、行方向又は列方向の複数の記憶素子間で共通な配線からの電圧を前記記憶素子に印加するときは導通する他の記憶素子である請求項 1 に記載の不揮発性半導体記憶装置。

【請求項 5】行方向又は列方向の複数の記憶素子間で共通な配線と前記記憶素子との接続および遮断をそれぞれ制御する複数の選択素子と、当該複数の選択素子の制御電極、前記複数の記憶素子の第 1 の制御電極、前記第 2 の制御電極、前記配線の印加電圧をそれぞれ制御して、書き込みを行う選択記憶素子に隣接した前記素子間半導体領域に反転層を形成し、当該反転層に、対応する導通状態の選択素子を介して前記所定電位を供給する一方で、第 1 の制御電極が前記選択記憶素子と共通接続された非選択な記憶素子に隣接した素子間半導体領域を、対応する選択素子を遮断し電気的に浮遊状態として空乏化し、当該素子間半導体領域に所定の書き込み阻止電位を誘起させるバイアス制御手段とを更に有する請求項 1 に記載の不揮発性半導体記憶装置。

【請求項 6】ビット線と共通電位線との間に、それぞれ選択素子を介して複数の記憶素子が直列接続されてなる記憶素子列を、行列状に複数配置させてメモリアレイが構成され、前記第 2 の制御電極が、前記記憶素子列の各記憶素子

2

間、記憶素子列両端の記憶素子と選択素子間に設けられている請求項 1 に記載の不揮発性半導体記憶装置。

【請求項 7】前記第 2 の制御電極は、前記第 1 の制御電極および前記電荷蓄積手段に対し絶縁膜を介して覆うプレート形状を有し、かつ、少なくとも行方向に並ぶ複数の前記記憶素子列間で共通に設けられている請求項 6 に記載の不揮発性半導体記憶装置。

【請求項 8】前記電荷蓄積手段は、少なくとも前記チャネル形成領域と対向する面内で離散化されている請求項 1 に記載の不揮発性半導体記憶装置。

【請求項 9】前記記憶素子は、前記チャネル形成領域上に、トンネル絶縁膜、窒化膜、トップ酸化膜を前記第 1 の制御電極との間に積層させてなり、前記電荷蓄積手段は、上記積層膜内で離散化して形成された電荷トラップである請求項 8 に記載の不揮発性半導体記憶装置。

【請求項 10】前記記憶素子は、前記チャネル形成領域上に、トンネル絶縁膜、窒化膜を前記第 1 の制御電極との間に積層させてなり、

前記電荷蓄積手段は、上記積層膜内で離散化して形成された電荷トラップである請求項 8 に記載の不揮発性半導体記憶装置。

【請求項 11】半導体のチャネル形成領域上に、電荷蓄積手段を含む絶縁膜を介して制御電極を積層させてなる複数の記憶素子を行列状に配置させてメモリセルアレイが構成され、前記制御電極にプログラム電圧を印加することにより前記電荷蓄積手段に電荷を電気的に注入し又は当該電荷蓄積手段から電荷を引き抜いて情報を記憶する不揮発性半導体記憶装置の書き込み方法であって、書き込みを行う前記記憶素子について、そのチャネル形成領域に隣接した半導体領域に少数キャリアによる反転層を形成し、当該反転層を上記プログラム電圧印加による電位との電位差により書き込みが行われる所定電位で保持して書き込みを行い、前記記憶素子と制御電極同士が接続された情報を書き込まない他の記憶素子について、そのチャネル形成領域に隣接した半導体領域に上記書き込み中に空乏層を形成し、書き込み阻止電位を誘起させる不揮発性半導体記憶装置の書き込み方法。

【請求項 12】前記電荷蓄積手段は、少なくとも前記チャネル形成領域と対向する面内で離散化されている請求項 11 に記載の不揮発性半導体記憶装置の書き込み方法。

【請求項 13】前記記憶素子は、前記チャネル形成領域上に、トンネル絶縁膜、窒化膜、トップ酸化膜を前記制御電極との間に積層させてなり、前記電荷蓄積手段は、上記積層膜内で離散化して形成された電荷トラップである請求項 12 に記載の不揮発性半導体記憶装置の書き込み方法。

【請求項 14】前記記憶素子は、前記チャネル形成領域

上に、トンネル絶縁膜、窒化膜を前記制御電極との間に積層させてなり、前記電荷蓄積手段は、上記積層膜内で離散化して形成された電荷トラップである請求項12に記載の不揮発性半導体記憶装置の書き込み方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、チャネル形成領域と制御電極との間の絶縁膜内に電荷蓄積手段を有し、当該電荷蓄積手段に電荷を注入し又は引き抜くことにより、電気的にデータのプログラムが可能な不揮発性半導体記憶装置とその書き込み方法に関する。特定的には、本発明は、記憶素子の書き込み時間が長い場合でも、当該記憶素子と制御電極同士が接続され書き込みを禁止したい他の記憶素子について、書き込みを禁止できる時間を長くして誤書き込みを有効に防止できる構成の不揮発性半導体記憶装置とその書き込み方法に関する。更に、本発明は、記憶素子間、記憶素子と選択素子間の半導体表面に第2の制御電極により空乏層を誘起して、その空乏層の電位を第2の制御電極の電位により制御することで、書き込み禁止をより確実に行うことができる不揮発性半導体記憶装置とその書き込み方法に関する。

【0002】

【従来の技術】従来、不揮発性半導体メモリ素子は、半導体のチャネル形成領域（素子のチャネルが形成される、例えば半導体基板又はウェル等の表面領域）と制御電極との間に介在させた絶縁膜中に電荷蓄積手段を設け、この電荷蓄積手段に存在する電荷の有無、或いは電荷量に応じてメモリ素子のしきい値（一般には、ゲートしきい値電圧）を変化させ、このしきい値を記憶データ信号の論理状態に対応させている。ここで、電荷を蓄積する電荷蓄積手段には、例えば、FG（フローティングゲート）等の単一導電層、ナノクリスタル等の平面的に離散化されて複数配置され互いに絶縁された小粒径導電体、窒化膜中或いは窒化膜と酸化膜界面に形成され空間的（平面方向および膜厚方向）に離散化された電荷トラップ等がある。この電荷トラップを含むゲート絶縁膜の積層構造の違いに応じて、FG型、ナノ結晶型、MONOS型、MNOS型等の様々な種類の不揮発性半導体メモリ素子が試作、提供されている。

【0003】かかるメモリ素子を多数配置させてメモリセルアレイが構成された不揮発性半導体記憶装置（不揮発性メモリ装置）では、多くの種類のメモリセル方式が提案されているが、その中でもっともセルサイズが小さく大容量化が可能なセル方式として、NAND型がある。NAND型不揮発性メモリ装置は、複数のメモリトランジスタを直列接続してNAND列と称されるメモリブロックを構成し、2個のNAND列で1個のビットコンタクトおよびソース線を共有することにより、1ビットあたりの実効的なセル面積の縮小を可能としたもので

ある。

【0004】図13は、従来のNAND型不揮発性メモリ装置のメモリセルアレイの基本構成を示す回路図である。

【0005】図13中、符号100はメモリセルアレイ、M11a~M1na, M11b~M1nb, M21a及びM21bはメモリトランジスタ、S11a, S12a, S11b, S12b, S21a, S22a, S21b, S22bは選択トランジスタ、BLa, BLbはビット線、WL11~WLn1及びWL21はワード線、SLはソース線、SG11及びSG21はビット線選択信号線、SG12及びSG22はソース線選択信号線、BCはビットコンタクトを示す。ストリングと称される繰り返し単位は、ビット線又はソース線に接続された2つの選択トランジスタ（選択ゲート）と、両選択トランジスタ間にn個（nは、例えば8, 16, 32等の数）のメモリトランジスタを直列接続させたNAND列とから構成されている。ビット線に接続された選択トランジスタS11a, S11b, S21a及びS21bは、ビット線選択信号線SG11又はSG21により制御され、ソース線に接続された選択トランジスタS12a, S12b, S22a及びS22bは、ソース線選択信号線SG12又はSG22により制御される。また、メモリトランジスタM11aとM11b, M12aとM12b, M13aとM13b, M1naとM1nbは、それぞれワード線WL11, WL12, WL13, WLn1により制御される。同様に、メモリトランジスタM21aとM21bは、ワード線WL21により制御される。

【0006】図14は、従来の不揮発性メモリ装置において、図13の左下に位置するNAND列を中心とした列方向断面図である。なお、ここでは、メモリトランジスタをFG（Floating Gate）型とした場合を例示する。図14において、符号2は例えばn型の半導体基板、4は例えばp型のウェル（pウェル）、24は層間絶縁層、24aは層間絶縁層24に開孔されたビットコンタクト孔を示す。ビットコンタクト孔24aは、これに埋め込まれた接続プラグとともに前記ビットコンタクトBCを構成する。各メモリトランジスタM11a~M1naは、pウェル4上に、トンネル絶縁膜40、フローティングゲートFG、ゲート間絶縁膜42、コントロールゲートCGが積層されて構成されている。各メモリトランジスタのコントロールゲートCGは、それぞれワード線WL11~WLn1を構成する。

【0007】選択トランジスタSG11, SG12, SG21は、メモリトランジスタと基本的には同じ積層構造であるが、これら選択トランジスタでは、メモリトランジスタにおいてフローティングゲートFGとなる層とコントロールゲートCGとなる層がゲート間絶縁膜42に設けられた接続孔を介して短絡されている。これによ

5

り、通常の単層ゲートと同じく、ゲート絶縁膜上のゲート電極層は全て同電位になっており、これによりビット線選択信号線 SG 1 1, SG 2 1、及び、ソース線選択信号線 SG 1 2 が構成されている。

【0008】このように配置されたゲート電極間のスペース領域に位置する p ウェル 4 の表面領域には、メモリトランジスタ及び選択トランジスタのソース・ドレイン不純物領域 6 c が形成されている。一方の選択トランジスタ SG 1 1, SG 2 1 のゲート電極の離間スペースに位置する p ウェル 4 の表面領域には、ビット方向の 2 つのストリング間で共通なドレイン不純物領域 6 a が形成されている。また、他方の選択トランジスタ SG 1 2 のゲート電極外側に位置する p ウェル 4 の表面領域には、ビット方向に隣接する他のストリング間で共通な前記ソース線 SL をなすソース不純物領域 6 b が形成されている。

【0009】図 1 4 では FG 型を示すが、電荷蓄積手段が平面的に離散化された不揮発性メモリ素子を用いた場合でも、ゲート絶縁膜構造が異なるほかは、従来の NAND 型では基本的には図 1 4 と同様である。p ウェル 4 とゲート電極（ワード線）との間に、MONOS 型ではトンネル絶縁膜、窒化膜及びトップ酸化膜を積層させ、MNOS 型ではトンネル絶縁膜と窒化膜を積層させている。また、ナノ結晶型では、p ウェル 4 上のトンネル絶縁膜上の絶縁膜内に小粒径導電体を互いに離散化させて埋め込んでいる。

【0010】つぎに、このような構成の NAND 型不揮発性メモリ装置において、メモリトランジスタがノーマリーオン状態と、ノーマリーオフ状態に対応させて 2 値情報が記憶されている場合を例に、一般的な動作説明を行う。読み出し動作では、読み出すセル（選択セル）が接続されたワード線（選択ワード線）とウェルを 0 V に電位固定し、全ての選択トランジスタと、選択ワード線以外のワード線（非選択ワード線）に接続されたメモリトランジスタとの全てが導通するような電圧 VRG を、全ての選択信号線と非選択ワード線に印加する。この電圧 VRG は、ウェルとの電位差のみでは、メモリトランジスタに書き込みと消去がされない大きさで、例えば 5 V ~ 7 V 程度である。この状態で、選択セルが接続されたビット線（選択ビット線）のみ正の電圧を印加すると、情報を読み出すセル以外の全てのメモリトランジスタは導通状態にあるため、選択セルのメモリトランジスタが、ノーマリーオンかノーマリーオフかによって、選択ビット線に電流が流れるか流れないかが決まる。この電流の有無を検出し、記憶データの論理状態 “1” 又は “0” を判定する。

【0011】消去動作は、通常、ブロック単位で行われ、選択ブロックの全ワード線に 0 V、非選択 NAND 列の全ワード線および基板又はウェルに高電圧 V_{pp} を印加する。その結果、選択ブロックのメモリトランジスタ

6

のみ、フローティングゲートから基板に電子が引き抜かれて、メモリトランジスタのしきい値電圧は負方向にシフトして、例えばノーマリーオンの消去状態（論理状態は例えば “1” に対応）になる。

【0012】一方、データのプログラム動作は、通常、選択するワード線に接続されたメモリトランジスタ括弧に、いわゆるページ単位で行われる。具体的には、ビット線側の選択トランジスタをオン、ソース線側の選択トランジスタをオフした状態で、選択ワード線に高電圧、非選択ワード線に非選択セルに書き込みは行われないがオンする程度の中間電圧（パス電圧）を印加する。このとき、プログラム（例えば “0” データを記憶）すべきメモリトランジスタが接続された選択ビット線に 0 V、プログラムを禁止（例えば “0” データを保持）すべきメモリトランジスタのみ接続された非選択ビット線に上記印加高電圧による電位との電位差により書き込みがされない程度に高い中間電位を設定しておく。その結果、プログラムすべき選択メモリトランジスタのみ、フローティングゲート中に電子が注入されて、選択メモリトランジスタのしきい値電圧は正方向にシフトして上記消去状態より高い、例えばノーマリーオフの書き込み状態になる。

【0013】この書き込み動作において、通常、高い負荷容量を有するビット線の充放電により動作時間が律束されるので、ビット線を駆動する電圧を低減して昇圧系回路の負担を軽減することを主な目的として、非選択 NAND 列のみ選択ゲートでビット線から切り離す技術が、特許公開公報平 6 - 9 7 4 5 5 号に記載されている。この技術では、非選択ビット線の印加電圧を選択ゲートの印加電圧からしきい値を引いた程度に低減して、非選択 NAND 列をビット線から切り離すこととしている。この結果、以後はパス電圧やプログラム電圧による自動昇圧により、非選択 NAND 列のチャネル電位（書き込み阻止電位）が設定されることから、上記文献に記載された技術は、現在、セルフブースト技術として広く知られている。

【0014】

【発明が解決しようとする課題】しかしながら、このセルフブースト技術を、例えば図 1 4 の断面図で基本構成を示した従来の不揮発性メモリ装置に適用した場合、以下に示す幾つかの問題がある。

【0015】不揮発性メモリでは、電源電圧の低減、セル微細化が進むにつれてトンネル絶縁膜を更に薄膜化しているが、特に FG 型ではストレスリークに起因した膜厚限界が存在することが指摘され、その膜厚限界は理論値でも 6 nm、現実には 8 nm 程度とされており（日経マイクロデバイス 1 月号及び 2 月号、1997 年 参照）、トンネル膜厚がスケールされないまま低電圧駆動が進むと、書き込み速度が遅くなる可能性がある。

第 1 の問題は、このような理由で書き込み速度が遅い場

7

合に、データを書き込んでいる選択セルと同じワード線に接続された非選択セルのメモリトランジスタが、誤書き込みされやすいことである。先に挙げた文献（公開公報）によれば、非選択NAND列を有効に書き込み禁止状態にしてしきい値シフトがない時間は、ある一定の上限があり、例えば10msecを越えとしきい値が上昇することがグラフで示されている。したがって、書き込み時間がこの上限を越えて長くなると、書き込みの最終段階で同じ選択ワード線に接続された非選択セルのメモリトランジスタに電荷が注入されて、多少なりとも書き込みされてしまう。

【0016】この第1の問題は、基本的にはMONOS型、MNOS型等においても同じであるが、MONOS型、MNOS型等では、電荷トラップが離散化されているのでFG型に比べゲート絶縁膜のスケリング性に優れるとされることから、プログラム電圧の低下に伴いゲート絶縁膜等のスケリングを行うことで、書き込み時の低電圧化に伴って書き込み速度が低下しにくく、この点ではFG型より有利である。

【0017】第2の課題として、素子の微細化に伴い減少化傾向にある書き込み状態と消去状態のしきい値のシフト量（Vthウインドウ幅）がプログラム電圧の低電圧化に適合しないことが挙げられる。すなわち、従来より低い電圧でプログラムされるように記憶素子のトンネル絶縁膜等が最適化されているNAND型不揮発性メモリにおいて、Vthウインドウ幅もある程度縮小化されないと、特に読み出し時にNAND列の非選択メモリトランジスタをオンさせる電圧を低くできず、しきい値のパラツキに応じて弱く書き込まれる非選択メモリトランジスタが生じやすくなる。この点でも、誤書き込みが発生しやすい傾向にある。

【0018】本発明は、このような実情に鑑みてなされ、プログラム電圧の低減に伴う非選択セルへの誤書き込みを有効に防止する構造の不揮発性半導体記憶装置を提供することを目的とする。また、本発明は、この不揮発性半導体記憶装置において好適に実施でき、プログラム電圧の低減に伴う非選択セルへの誤書き込みを有効に防止できる不揮発性半導体記憶装置の書き込み方法を提供することを他の目的とする。

【0019】

【課題を解決するための手段】上述した従来技術の問題点を解決し、上記目的を達成するために、本発明の不揮発性半導体記憶装置では、例えば従来構成のNAND列等、トランジスタが直列接続されるようなセル構成において、そのトランジスタ間に共通な不純物添加領域（ソース・ドレイン領域）がない構成とし、当該半導体領域を、特に選択メモリトランジスタと同一ワード線に接続された非選択メモリトランジスタ側で容量結合によって空乏化させ、これにより書き込み阻止電位を誘起させることとした。すなわち、本発明の不揮発性半導体記憶装

8

置は、半導体のチャネル形成領域上に電荷蓄積手段を含む絶縁膜を介して積層された第1の制御電極に電圧を印加し、前記電荷蓄積手段に対し電荷を電気的に注入し又は当該電荷蓄積手段から電荷を引き抜くことにより情報を記憶する記憶素子と、前記電荷の注入又は引き抜きの際に前記チャネル形成領域に所定の電位を伝達する素子とを有する不揮発性半導体記憶装置であって、前記記憶素子と前記所定電位を伝達する素子間の半導体領域と容量結合し、当該半導体領域に対し素子間チャネルまたは空乏層の形成と電位の制御を行う第2の制御電極が、当該記憶素子と所定電位を伝達する素子との間の半導体領域上に絶縁膜を介して設けられている。この所定電位を伝達する素子は、例えばNAND型等の他の記憶素子であってもよく、また、行方向又は列方向の複数の記憶素子間で共通な配線と前記記憶素子との接続と遮断を制御する選択素子であってもよい。

【0020】好ましくは、前記第2の制御電極は、前記第1の制御電極および前記電荷蓄積手段に対し絶縁膜を介して覆うプレート形状を有し、かつ、少なくとも行方向に並ぶ複数の前記記憶素子列間で共通に設けられている。また、好ましくは、行方向又は列方向の複数の記憶素子間で共通な配線と前記記憶素子との接続および遮断をそれぞれ制御する複数の選択素子と、当該複数の選択素子の制御電極、前記複数の記憶素子の第1の制御電極、第2の制御電極、前記配線の印加電圧をそれぞれ制御して、書き込みを行う選択記憶素子に隣接した前記素子間半導体領域に反転層を形成し、当該反転層に、対応する導通状態の選択素子を介して前記所定電位を供給する一方で、第1の制御電極が前記選択記憶素子と共通接続された非選択記憶素子に隣接した素子間半導体領域を、対応する選択素子を遮断し電気的に浮遊状態として空乏化し、当該素子間半導体領域に所定の書き込み阻止電位を誘起させるバイアス制御手段とを、更に有する。

【0021】本発明の不揮発性半導体記憶装置の書き込み方法は、半導体のチャネル形成領域上に、電荷蓄積手段を含む絶縁膜を介して制御電極を積層させてなる複数の記憶素子を行列状に配置させてメモリセルアレイが構成され、前記制御電極にプログラム電圧を印加することにより前記電荷蓄積手段に電荷を電気的に注入し又は当該電荷蓄積手段から電荷を引き抜いて情報を記憶する不揮発性半導体記憶装置の書き込み方法であって、書き込みを行う前記記憶素子について、そのチャネル形成領域に隣接した半導体領域に少数キャリアによる反転層を形成し、当該反転層を上記プログラム電圧印加による電位との電位差により書き込みが行われる所定電位で保持して書き込みを行い、前記記憶素子と制御電極同士が接続された情報を書き込まない他の記憶素子について、そのチャネル形成領域に隣接した半導体領域に上記書き込み中に空乏層を形成し、書き込み阻止電位を誘起させる。

【0022】このような不揮発性半導体記憶装置及びそ

9

の書き込み方法では、それぞれ所定電位が印加された前記第1及び第2の制御電極の容量結合により、チャネル形成領域と素子間半導体領域に少数キャリアが引き寄せられて、メモリトランジスタを単数又は複数個含むトランジスタ列ごとに反転層が形成される。ところが、トランジスタ列が複数ある場合、その素子間の半導体領域の表面側部分が配線に電気的に接続されて、ある電位が供給されている場合と、電位を供給する配線と切り離されている場合とでは、隣接する領域にチャネルを形成する際に時間差が生じる。このような制御では、例えばセルフブースト技術を用いて書き込みを禁止するNAND型の場合で説明すると、選択ブロックのNAND列端に接続された選択トランジスタがオンしてビット線電位（例えば、0V）が供給されているので、ビット線から少数キャリアが急速に供給されて、この選択ブロックのNAND列は、チャネルが早く形成される。一方、非選択ブロックでは、ビット線側の選択トランジスタが遮断するので、表面にチャネル形成領域を有するウェル等は例えば接地電位に固定されているものの、チャネル形成領域の表面領域が電位固定されていないことから、少数キャリアの急速な供給源がない。このため、ビット線から所定電位（例えば、接地電位）が供給されている場合と比較してチャネルの形成が遅く、空乏化した状態が長く保持される。このとき非選択ブロックの素子間半導体領域は、第2の制御電極との容量結合によって電位がプルアップされ、この結果、当該素子間半導体領域、ひいては隣接するチャネル形成領域に所定の書き込み阻止電位を誘起することが可能となる。

【0023】従来では、非選択ブロックはビット線等の配線に接続されなくても、NAND列の途中で所定間隔でチャネル形成領域を挟んで交互に、チャネル形成領域と逆導電型の不純物拡散層（ソース・ドレイン領域）が設けられているので、これが少数キャリアの供給源として機能し、選択ブロックのチャネル形成より遅いが、それでも比較的速い段階でチャネルが形成されてしまう。また、チャネルで連続した素子間の電位差は少なくなり、書き込み禁止をしたい記憶素子のチャネル形成領域の電位を隣接する記憶素子のチャネル形成領域の電位よりその絶対値を大きく設定して、書き込み禁止を確実にすることは難しかった。

【0024】これに対し、本発明の不揮発性半導体記憶装置では、このようなキャリア供給源として作用する不純物拡散層がなく、その部分に隣接するチャネル形成領域でのチャネル（反転層）の形成を従来より遅くでき、この結果、書き込み禁止時間が長くなる。つまり、誤書き込みに対する時間的なマージンが拡大する。また、このチャネル形成を遅くできることは、ある時刻でみると、トンネル絶縁膜にかかる電位差に関し、誤書き込みが生じる限界値との電圧マージンが拡大することを意味する。さらに、記憶素子間、記憶素子と選択素子間の半

10

導体領域表面に空乏層を誘起し、その表面電位を第2の制御電極の電位で制御できるので、書き込み禁止をしたい記憶素子のチャネル形成領域の表面電位の絶対値を大きく設定することが可能となり、書き込み禁止を確実にすることができる。

【0025】とくに、電荷蓄積手段が平面内で離散化されている場合、上記電圧マージンをさらに大きくできる。FG型では、第2の制御電極の印加電圧を大きくすると、空乏層（素子間半導体領域）の電位が上昇するが容量結合によって電荷蓄積手段（浮遊ゲート）電位もある程度上昇してしまう。これに対し、電荷蓄積手段が平面内で離散化されている場合、これと第2の制御電極とは殆ど容量結合しないので、誤書き込みを引き起こすトンネル絶縁膜にかかる電位差を急速に低下させることができる。

【0026】

【発明の実施の形態】本発明は、記憶素子が、その制御電極の容量結合を利用して電荷蓄積手段（例えば、フローティングゲート、電荷トラップ等）に対し電荷を電気的に注入し又は引き抜くことを基本動作とし、その記憶素子が、他の記憶素子又は選択トランジスタ等と直列接続されるセル構成を有する不揮発性半導体記憶装置に広く適用される。このような不揮発性半導体記憶装置であれば、記憶素子の種類に限定はなく、通常のFG（Floating Gate）型のほか、MNOS（Metal-Nitride-Oxide Semiconductor）、MONOS（Metal-Oxide-Nitride-Oxide Semiconductor）、更にはナノ結晶型等に適用可能である。また、メモリセル方式に限定はなく、NAND型のほか、NOR型（ビット線、ソース線が階層化されたもの、例えばDINOR型或いはAND型を含む）に適用可能である。以下、MONOS、MNOS型及びFG型FETを記憶素子として有するNAND型不揮発性メモリ装置を例として、本発明に係る不揮発性半導体記憶装置及びその製造方法の実施形態を、図面にもとづいて詳細に説明する。

【0027】第1実施形態

本実施形態は、MONOS型に関するものである。図1は、本実施形態に係るNAND型不揮発性メモリ装置のメモリセルアレイの回路構成図である。また、図2は、本実施形態に係るNAND列（図1の左下のNAND列に対応）の断面構造図である。

【0028】図1において、各ビット線BLa、BLbには、それぞれ2本のトランジスタ列（ストリング）が接続されている。ビット線BLaに接続された第1のストリングは、ビット線BLaに接続された選択トランジスタS11aと、ソース線SLに接続された選択トランジスタS12aと、両選択トランジスタ間に直列接続されたn個のメモリトランジスタM11a～M1naとからなる。ビット線BLaに接続された第2のストリングは、ビット線BLaに接続された選択トランジスタS2

11

1 a と、ソース線 S L に接続された選択トランジスタ S 2 2 a と、両選択トランジスタ間に直列接続された n 個のメモリトランジスタ M 2 1 a … とからなる。同様に、ビット線 B L b にも 2 本のストリングが接続されており、その第 1 のストリングは、ビット線 B L b に接続された選択トランジスタ S 1 1 b と、ソース線 S L に接続された選択トランジスタ S 1 2 b と、両選択トランジスタ間に直列接続された n 個のメモリトランジスタ M 1 1 b ~ M 1 n b とからなる。ビット線 B L b に接続された第 2 のストリングは、ビット線 B L b に接続された選択トランジスタ S 2 1 b と、ソース線 S L に接続された選択トランジスタ S 2 2 b と、両選択トランジスタ間に直列接続された n 個のメモリトランジスタ M 2 1 b … とからなる。

【0029】選択トランジスタ S 1 1 a および S 1 1 b は、ともにビット線選択信号線 S G 1 1 により制御され、選択トランジスタ S 1 2 a および S 1 2 b は共にソース線選択信号線 S G 1 2 により制御される。同様に、選択トランジスタ S 2 1 a および S 2 1 b は、ともにビット線選択信号線 S G 2 1 により制御され、選択トランジスタ S 2 2 a および S 2 2 b は共にソース線選択信号線 S G 2 2 により制御される。また、メモリトランジスタ M 1 1 a と M 1 1 b, M 1 2 a と M 1 2 b, M 1 3 a と M 1 3 b, …, M 1 n a と M 1 n b は、それぞれワード線 W L 1 1, W L 1 2, W L 1 3, …, W L 1 n により制御される。同様に、メモリトランジスタ M 2 1 a と M 2 1 b は、ワード線 W L 2 1 により制御される。

【0030】このように相互接続される各ストリングにおいて、図 2 の断面に示すように、例えば n 型の半導体基板 2 内の表面側に、例えば p 型のウェル (p ウェル 4) が形成され、当該 p ウェルの表面側にトランジスタ列が配置させている。

【0031】メモリトランジスタ M 1 1 a ~ M 1 n a は、その p ウェル 4 上に、トンネル絶縁膜 8、窒化膜 10、トップ酸化膜 12、からなるゲート絶縁膜が構成されている。また、ゲート絶縁膜上に、ポリシリコン層 16 と、その上の高融点金属シリサイド層 18 からなるワード線 W L 1 1 ~ W L 1 n が積層されている。このワード線は、本発明の“第 1 の制御電極”に該当する。また、ポリシリコン層と高融点金属シリサイド層からなる第 1 の制御電極下方の p ウェル表面部分が本発明における“チャネル形成領域”、当該第 1 の制御電極間の p ウェル表面部分が本発明における“素子間半導体領域”に該当する。

【0032】選択トランジスタ S 1 1 a, S 1 2 a 等は、通常の MOSFET で構成される。したがって、そのゲート絶縁膜 14 は、シリコン酸化膜のみで構成されている。選択トランジスタ S 1 1 a, S 1 2 a の各ゲート電極層は、それぞれビット線選択信号線 S G 1 1, ソース線選択信号線 S G 1 2 を構成する。

12

【0033】本実施形態の不揮発性メモリの構造的な特徴の一つは、メモリトランジスタ間、メモリトランジスタと選択トランジスタ間の p ウェル 4 の表面領域 (素子間半導体領域) に、この p ウェル 4 と逆導電型の不純物領域 (従来例を示す図 14 におけるソース・ドレイン領域 6 c) が形成されていないことである。他の逆導電型の不純物領域、即ちビット方向の一方側に隣接する他のストリングと共通な逆導電型のドレイン不純物領域 6 a と、他方の隣接ストリング間で共通な逆導電型のソース不純物領域 6 b (ソース線 S L) は、従来と同様に設けられている。

【0034】本実施形態の不揮発性メモリの構造的な他の特徴は、トランジスタ列上に例えば酸化シリコンからなる層間絶縁層 20 を介して、第 2 の制御電極 22 が設けられていることである。この第 2 の制御電極 22 は、例えば各トランジスタ列ごとに、その一方の選択トランジスタの上方から他方の選択トランジスタの上方にかけて一枚のプレート状に形成されている。そして、第 2 の制御電極 22 は、各メモリトランジスタ間、或いはメモリトランジスタと選択トランジスタ間の素子間半導体領域に対し層間絶縁層 20 を介して近接しており、当該制御電極 22 の印加電圧に応じて、その下方の p ウェル 4 表面部分 (素子間半導体領域) において素子間チャネルまたは空乏層の形成ならびに電位を制御する。

【0035】第 2 の制御電極 22 上を含む全面に、比較的に厚い層間絶縁層 24 が成膜され、ビットコンタクト B C は、この層間絶縁層 24 について前記ドレイン不純物領域 6 a 上で開孔されたコンタクト孔 24 a 内を、例えば T i / T i N 等の密着層を介在させて W 等の金属プラグで埋め込むことにより形成されている。ビットコンタクト B C によりトランジスタ列と接続するかたちで、ビット線 B L a が層間絶縁層 24 上に配線されている。ビット線 B L a は、特に図示しないが、通常、例えば A l 等の主配線層の上下を、反射防止層 (又は保護層) とバリアメタルで挟んだ 3 層構造を有している。

【0036】このような構成の N A N D 型不揮発性メモリ装置では、メモリトランジスタと隣接する他のトランジスタ間に不純物領域がないので、後述するように誤書き込みが有効に防止される利点がある。

【0037】つぎに、製造方法について説明する。図 3 ~ 図 6 は、上述した構成の不揮発性メモリ装置の各製造過程を示す断面図である。

【0038】この図 3 に先立って、まず、用意した基板 (n 型半導体基板 2) の表面部分に、例えば L O C O S 又はトレンチによる素子分離領域を形成する。この素子分離領域は、断面図には現れないが、各トランジスタ列と交互にビット方向に長い平行ストライプ状に形成され、これにより主に行方向の素子分離が達成される。

【0039】ついで、図 3 (a) に示すように、基板内の表面部分に p ウェル 4 をイオン注入法により形成す

10

20

30

40

50

13

る。なお、このpウェル4は、エピタキシャル成長層、基板（石英、サファイア基板等の半導体以外も可）上に絶縁層を介して形成したSOI層等で代替できる。pウェル4表面を例えば熱酸化することにより、トンネル絶縁膜8を形成する。トンネル絶縁膜8は、酸化シリコンのほか、熱酸化シリコンを熱窒化処理することにより、その表面部分或いは全ての熱酸化シリコンを窒化酸化シリコンとしてもよい。トンネル絶縁膜8上に、例えば窒化シリコンからなる窒化膜10をCVD法等により成膜する。

【0040】つぎに、図4（b）に示すように、メモリトランジスタ列が形成される部分にのみ被膜するレジストパターンRPを形成し、これをマスクに下地の窒化膜10をエッチングし、選択トランジスタが形成される部分の窒化膜を除去する。このエッチングの際に、トンネル絶縁膜8は少なくとも一部残して、基板側にエッチングダメージが導入されないようにすることが望ましい。

【0041】続く図4（c）では、熱酸化を行う。先の工程でトンネル絶縁膜8を少なくとも一部残した場合、熱酸化の前処理等のウェット処理で完全に除去することが望ましい。この熱酸化により、ウェル4がほぼ表出した選択トランジスタ部には、選択トランジスタのゲート絶縁膜14となる熱酸化膜が厚く形成される一方、窒化膜10上では、メモリトランジスタのトップ酸化膜12となる熱酸化膜が薄く形成される。このとき、窒化膜10が膜減りして最終的な所望の窒化膜厚が得られる。

【0042】図5（d）では、熱酸化膜12、14上に、不純物が導入されて導電化されたポリシリコン層16と、高融点金属シリサイド層18を、通常のポリサイド形成法により積層させる。

【0043】続く図5（e）では、レジストパターンRPを用いて、これをマスクに下地の積層膜をエッチングし、平行ストライプ状のワード線WL11～WL1n等と、選択信号線SG11、SG12等を同時形成する。この図では、シリコン層が表出するまで全てエッチオフしているが、基板側へのダメージ導入を抑制するには、全ての領域に最下層の酸化膜を部分的に残し、後でウェットエッチングにより除去してもよい。

【0044】つぎに、レジストパターンRPを除去後、図6（f）に示すように、例えば酸化シリコン等の第1の層間絶縁層20を全面に薄く成膜する。この成膜は、例えば熱酸化又はCVDの何れかの方法によって形成する。ワード線等がポリシリコンでなる場合、熱酸化法が採用できる。ワード線等がポリサイドの場合は、CVD法による。その後、第2の制御電極22となる導電層を全面に、しかもゲート電極間を埋め込むように成膜し、これを図示せぬレジストパターンをマスクとしてエッチング加工する。これにより形成された第2の制御電極22は、素子間チャネルまたは空乏層の形成ならびに電位を制御するために、少なくともメモリトランジスタの電

14

極間を覆っている必要がある。第2の制御電極はトランジスタ電極間のスペース毎に分離された埋め込みライン状に形成し、図示しない箇所でも電気的に接続させる構成でもよい。この第2の制御電極22、又はその形成マスクのレジストパターンを付けたままで、これをマスクにイオン注入を行い、第1の層間絶縁層20をスルー膜としてpウェル4の表面部分にドレイン不純物領域6aと、ソース線SLとなるソース不純物領域6bを同時形成する。

10 【0045】その後は、図2に示す如く、厚い第2の層間絶縁層24を堆積して平坦化し、ビットコンタクト孔24aの開孔、接続プラグの埋め込み、ビット線の配線等を行うことによって、当該不揮発性メモリ装置を完成させる。

【0046】つぎに、このような構成のNAND型不揮発性メモリ装置において、メモリトランジスタのデータ書き込み、消去及び読み出しの動作説明を行う。このメモリトランジスタは2値情報のほかに、多値情報を記憶する場合も本発明は適用できるが、ここでは、ノーマリーオンとノーマリーオフの2つの状態間での動作説明を行う。なお、多値の場合は、プログラム時或いは読み出し時のワード線電圧又はビット線を段階的に、例えば正方向にシフトさせて行うことから、基本的な動作は同じである。

【0047】まず、読み出し動作では、読み出すセル（選択セル）が接続されたワード線（選択ワード線）とウェルを0Vに電位固定し、全ての選択トランジスタと、選択ワード線以外のワード線（非選択ワード線）に接続されたメモリトランジスタとの全てが導通するような電圧VRGを、全ての選択信号線と非選択ワード線に印加する。この電圧VRGは、ウェルとの電位差のみでは、メモリトランジスタに書き込みと消去がされない大きさである。また、前記第2の制御電極に印加する電圧は、読み出すセルを含むNAND列において、その素子間チャネルが形成される電圧が選択される。この状態で、選択セルが接続されたビット線（選択ビット線）のみ正の電圧を印加すると、情報を読み出すセル以外の全てのメモリトランジスタは導通状態にあるため、選択セルのメモリトランジスタが、ノーマリーオンかノーマリーオフかによって、選択ビット線に電流が流れるか流れないかが決まる。この電流の有無を検出し、記憶データの論理状態“1”又は“0”を判定する。

【0048】消去動作は、従来と同様に、ブロック単位又はメモリセルアレイ一括して行われ、全ての選択トランジスタをオフした状態で、消去単位的全ワード線に0V、非選択NAND列の全ワード線および基板又はウェルに高電圧V_{pp}を印加する。その結果、選択ブロックのメモリトランジスタのみ、基板側から正孔が窒化膜および窒化膜と酸化膜の界面付近に存在する捕獲準位（電荷トラップ）に注入され、メモリトランジスタのしきい値

電圧は負方向にシフトして、例えばノーマリーオンの消去状態（論理状態は例えば“1”に対応）になる。

【0049】一方、データのプログラム動作では、ワード線に接続されたメモリトランジスタ一括に、いわゆるページ単位で行われる。具体的には、通常、選択ブロック行のソース側に近いページから書き込みを行うが、まず、ソース線側の選択トランジスタをオフした状態で、選択ビット線に例えば0V、ビット線側の選択トランジスタのゲート（ビット線選択信号線）および非選択ビット線に例えば電源電圧 V_{DD} 程度（例えば5V程度）を印加しておく。そして、非選択ワード線には、非選択セルが導通する程度の電圧、選択ワード線には高いプログラム電圧 V_{PP} を印加する。すると、選択ブロックのNAND列端に接続された選択トランジスタがオンしてビット線電位（例えば、0V）が供給されているので、ビット線から少数キャリアが急速に供給されて、この選択ブロックのNAND列は、チャンネルが早く形成される。その結果、プログラムすべき選択メモリトランジスタのみ、窒化膜中および窒化膜と酸化膜との界面付近に存在する捕獲準位に電子が注入されて、選択メモリトランジスタのしきい値電圧は正方向にシフトして上記消去状態より高い、例えばノーマリーオフの書き込み状態になる。

【0050】一方、非選択ブロックでは、そのビット線側の選択トランジスタがビット線選択信号線の印加電圧から選択ゲートのしきい値電圧を引いた僅かな電位上昇で導通状態から遮断状態に推移するので、表面にチャンネル形成領域を有するウェル自体は接地電位に固定されているものの、チャンネル形成領域端の表面領域が電位固定されていないことから、また、従来のように素子間に共通なソース・ドレイン不純物領域がなく、少数キャリアの急速な供給源がない。このため、ビット線から所定電位（例えば、接地電位）が供給されている選択NAND列側と比較してチャンネルの形成が遅く、空乏化した状態が長く（一定時間）保持される。そして、この空乏層が形成されている間に、非選択NAND列の素子間半導体領域はその電位が第2の制御電極22との容量結合によってブルアップされ、この結果、当該素子間半導体領域、ひいては隣接するチャンネル形成領域に所定の書き込み阻止電位が誘起され、誤書き込み防止が達成される。

【0051】このように本実施形態では、従来でソース・ドレイン不純物領域が形成されていた素子間半導体領域を第2の制御電極22との容量結合により空乏化した状態でセルフブーストすることから、この間にプログラム電圧 V_{PP} の印加パルス（書き込みパルス）が終了する時間的な余裕が生じ、即ち書き込み禁止時間が長くなる。結果として、誤書き込みに対する時間的なマージンが拡大する。また、このチャンネル形成を遅くできることは、ある時刻でみると、トンネル絶縁膜にかかる電位差に関し、誤書き込みが生じる限界値との電圧マージンが拡大することを意味する。とくに、本例では電荷蓄積手

段が平面内で離散化されているので、上記電圧マージンをさらに大きくできる。すなわち、FG型では第2の制御電極の印加電圧を大きくすると、空乏層（素子間半導体領域）の電位が上昇するが第2の制御電極との容量結合によって電荷蓄積手段（浮遊ゲート）電位もある程度上昇してしまうが、本例では電荷蓄積手段が平面内で離散化されているので、これと第2の制御電極とは殆ど容量結合しないので、誤書き込みを引き起こすトンネル絶縁膜にかかる電位差を急速に低下させることができる。

10 【0052】不揮発性メモリ装置では、プログラム動作が、一般に、書き込みパルスによるプログラムによって所望のしきい値が得られたかを検証する一種の読み出し動作を繰り返しながら実行される。すなわち、プログラム後の検証において所望のしきい値が得られると、そのメモリトランジスタについては、セルフブースト動作により書き込み禁止状態を設定しながら次のプログラムを行い、再度しきい値を検証する。これをワード線に接続されたメモリトランジスタ全てのしきい値について、所望の値が得られるまで繰り返す。

20 【0053】本実施形態の不揮発性メモリ装置及びその書き込み方法では、書き込み禁止時間を長くできることによって書き込みパルス幅を従来より大きくした制御が可能であり、特に最終的に制御するしきい値の許容幅が狭い多値メモリ技術において、書き込みパルス幅の制御の自由度が増大する利点がある。また、プログラム電圧が低電圧化された場合に、素子のスケールが低電圧化に適合しない等の理由から書き込み時間が遅くなる場合でも、有効に誤書き込みを防止できる。さらに、プログラム電圧が低電圧化傾向にあって、これに V_{th} ウインドウ幅の縮小化が追いつかない場合、読み出し時の非選択メモリトランジスタを全てオンさせるには、そのゲートに印加されるバス電圧の低下をなかなか進めることができないが、本発明によりチャンネル形成がされにくいことは、従来では弱い書き込み状態となる高いバス電圧のままでも誤書き込みが起きにくい作用をとともない、その意味でも、誤書き込みを有効に防止できる。

【0054】第2実施形態

本実施形態は、MNOS型不揮発性メモリ装置についてである。本実施形態の不揮発性メモリ装置は、そのメモリトランジスタを第1実施形態のMONOS型からMNOS型に変更したこと以外、図1に示す基本的なセルアレイの回路構成、動作は第1実施形態と変わらない。したがって、ここでは、NAND列の断面構造と製造方法のみを説明し、他は省略する。また、断面構成において重複する構成は、同じ符号を付して詳しい説明はしない。

【0055】図7は、本発明の第2実施形態に係る不揮発性メモリ装置において、そのNAND列のビット方向断面図である。本実施形態のメモリトランジスタM11a～M1naは、そのゲート絶縁膜が、pウェル4上に

17

トンネル絶縁膜30、窒化膜32を積層させてなる。また、選択トランジスタのゲート絶縁膜34は、第1実施形態と同様、酸化シリコンの膜のみから構成させてもよいが、ここでは酸化シリコン膜と窒化シリコン膜との積層膜としている。この場合、下層のシリコン酸化膜厚が例えば10nm程度の比較的に厚いことからメモリ素子として機能しない。このNAND型不揮発性メモリ装置は、第1実施形態と同様、メモリトランジスタと隣接する他のトランジスタ間に不純物領域がないので、誤書き込みが有効に防止される利点がある。

【0056】図8～図11は、上述した構成の不揮発性メモリ装置の各製造過程を示す断面図である。まず、第1実施形態と同様にして、用意した基板2に素子分離領域と、pウェル4を形成し、pウェル4表面を例えば熱酸化することにより、10nmほどの比較的に厚いゲート酸化膜5を形成する(図8(a))。

【0057】図9(b)に示すように、このゲート酸化膜5上にレジストパターンRPを形成し、これをマスクとしたエッチングにより、メモリトランジスタ部のゲート酸化膜5を除去する。続く図9(c)では、熱酸化法により、メモリトランジスタのトンネル絶縁膜30を形成する。このとき、ゲート酸化膜5は僅かに酸化膜厚が増大する。なお、トンネル絶縁膜30は、一旦熱酸化シリコン膜を形成した後、少なくとも表面部分を熱窒化してもよい。また、トンネル絶縁膜30上に、窒化膜32を例えばCVD法により数十nmと比較的に厚く堆積する。

【0058】以後は、第1実施形態と同様に、ポリシリコン層16と高融点金属シリサイド層18とからなるポリサイドを形成し(図10(d))、レジストパターンRPを用いて、これをマスクに下地の積層膜をエッチングしてワード線WL11～WL1n等と選択信号線SG11、SG12等を同時形成する(図10(e))。また、レジストパターンRPを除去後、第1の層間絶縁層20を薄く形成し(図11(f))、この上に第2の制御電極22を形成した後(図11(g))、第2の層間絶縁層24の堆積、ビットコンタクトBCの形成、ビット線BLaの配線等の諸工程を経て、当該不揮発性メモリを完成させる。

【0059】本実施形態の不揮発性メモリ装置においても、第1実施形態と同様な利点を有する。すなわち、従来のようにNAND列内のpウェル内表面側に、これと逆導電型の不純物領域を有しないので、セルフブーストにより非選択NAND列の書き込み禁止において、その書き込み禁止時間を長くできる。そのため、プログラム電圧の低電圧化により書き込み速度が遅い場合、この低電圧化に伴ってVthウィンドウ幅の縮小化が追いつかない場合等にあって、誤書き込みが有効に防止できる。また、書き込みパルス幅の制御の自由度が増大する。

【0060】第3実施形態

18

本実施形態は、FG型不揮発性メモリ装置についてである。本実施形態の不揮発性メモリ装置は、そのメモリトランジスタを第1実施形態のMONOS型からFGに変更したこと以外、図1に示す基本的なセルアレイの回路構成、動作は第1実施形態と変わらない。

【0061】図12は、本第3実施形態に係るFG型不揮発性メモリ装置において、そのNAND列のビット方向断面図である。この図12に示されるように、トンネル絶縁膜8と中間絶縁膜40(第1実施形態のトップ酸化膜に相当)との間に、第1実施形態の窒化膜に代えて、例えばポリシリコンからなるフローティングゲート42が介在する。中間絶縁膜40は、第1実施形態のトップ酸化膜と同様、酸化シリコン或いはONO膜等で構成される。同じ符号を付した他の構成は、第1実施形態と同じである。

【0062】このような構成のFG型不揮発性メモリ装置の製造方法は、第1実施形態における図3で窒化膜10でなくポリシリコン膜を成膜し、これを図4(b)でパターンニングすること、図4(c)でトップ酸化膜12と同様にして、中間絶縁膜を形成すること、図5

(e)の電極加工時に当該中間絶縁膜とポリシリコン膜を同時にパターンニングすること以外、第1実施形態の製造方法を示す図3～図6と同様に行うことができる。

【0063】本実施形態の不揮発性メモリ装置においても、第1実施形態と同様な利点を有する。すなわち、従来のように素子間に不純物拡散領域を有しないので、セルフブーストによる書き込み禁止において書き込み禁止時間を長くでき、非選択NAND列の誤書き込みが有効に防止でき、また書き込みパルス幅の制御の自由度が増大するといった数々の利点を有する。

【0064】

【発明の効果】本発明に係る不揮発性半導体記憶装置及びその書き込み方法によれば、プログラム電圧の低下にともなう誤書き込みを有効に防止し、遅い書き込み制御、例えば書き込みパルス幅を長くするといった場合の自由度が増大する。よって、微細化、低電圧書き込み化を推進した場合でも信頼性及び特性が優れた不揮発性半導体記憶装置と、その書き込み方法を提供することが可能となる。

【図面の簡単な説明】

【図1】本発明の実施形態に係るNAND型不揮発性メモリ装置のメモリアレイの構成を示す回路図である。

【図2】本発明の第1実施形態に係り、図1の左下に位置するストリングを中心としたビット方向の断面図である。

【図3】図2のNAND型不揮発性メモリ装置の各製造過程を示す断面図であり、ゲート絶縁膜を構成する窒化膜の成膜までを示す。

【図4】図3に続く同断面図であり、ゲート絶縁膜を構成するトップ酸化膜の成膜までを示す。

【図5】図4に続く同断面図であり、ワード線及び選択信号線とゲート絶縁膜のエッチング加工までを示す。

【図6】図5に続く同断面図であり、第2の制御電極形成までを示す。

【図7】本発明の第2実施形態に係り、図1の左下に位置するストリングを中心としたビット方向の断面図である。

【図8】図7のNAND型不揮発性メモリ装置の各製造過程を示す断面図であり、選択トランジスタのゲート絶縁膜となる酸化膜形成までを示す。

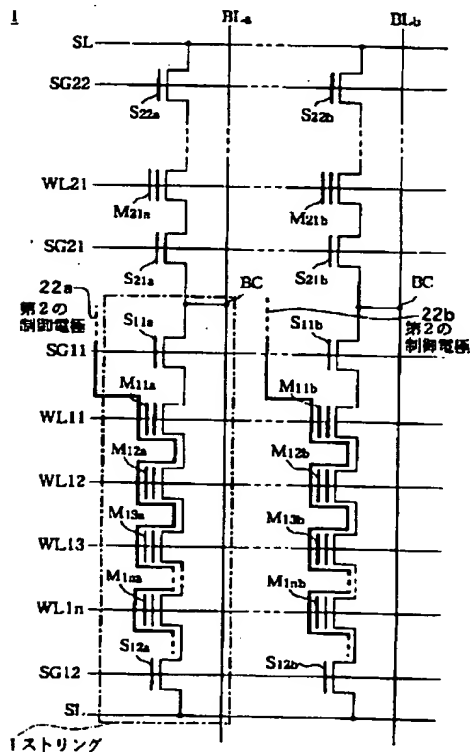
【図9】図8に続く同断面図であり、ゲート絶縁膜を構成する窒化膜の成膜までを示す。

【図10】図9に続く同断面図であり、ワード線及び選択信号線とゲート絶縁膜のエッチング加工までを示す。

【図11】図10に続く同断面図であり、第2の制御電極形成までを示す。

【図12】本発明の第3実施形態に係り、図1の左下に位置するストリングを中心としたビット方向の断面図で *

【図1】



*ある。

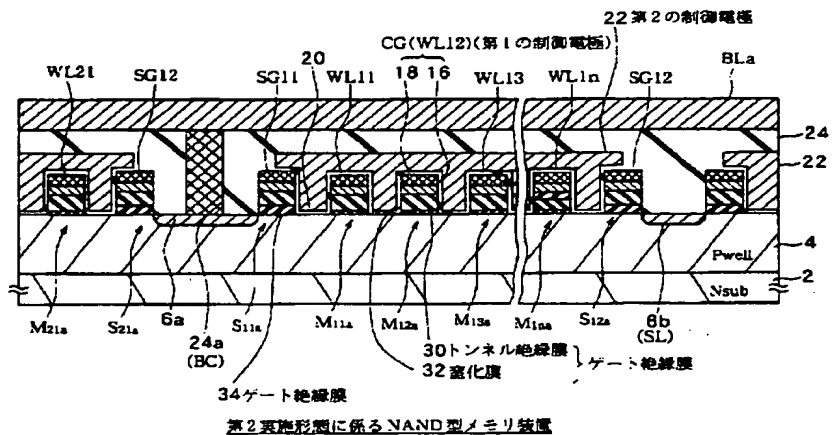
【図13】従来のNAND型不揮発性メモリ装置のメモリアレイの構成を示す回路図である。

【図14】従来例に係る図1の左下に位置するストリングを中心としたビット方向の断面図である。

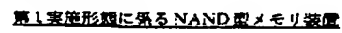
【符号の説明】

1…メモリアレイ、2…半導体基板、4…pウェル、5…酸化膜、6a~6c…不純物領域、8, 30…トンネル絶縁膜、10, 32…窒化膜、12…トップ酸化膜、14, 34…ゲート絶縁膜、16…ポリサイド層、18…高融点金属シリサイド層、20…第1の層間絶縁層（絶縁膜）、22…第2の制御電極、24…第2の層間絶縁層、40…中間絶縁膜、42, FG…フローティングゲート、M11等…メモリトランジスタ、ST11a, ST21a等…選択トランジスタ、CG…コントロールゲート、SG11, SG12等…選択信号線、BLa等…ビット線、WL11等…ワード線、BC…ビットコンタクト。

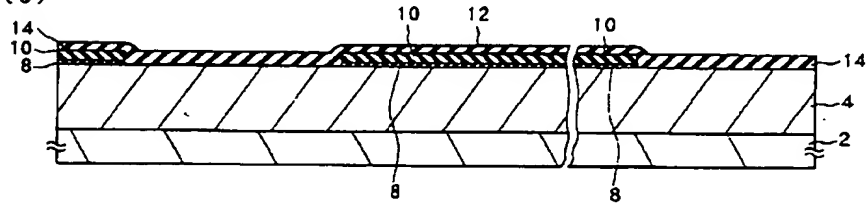
【図7】



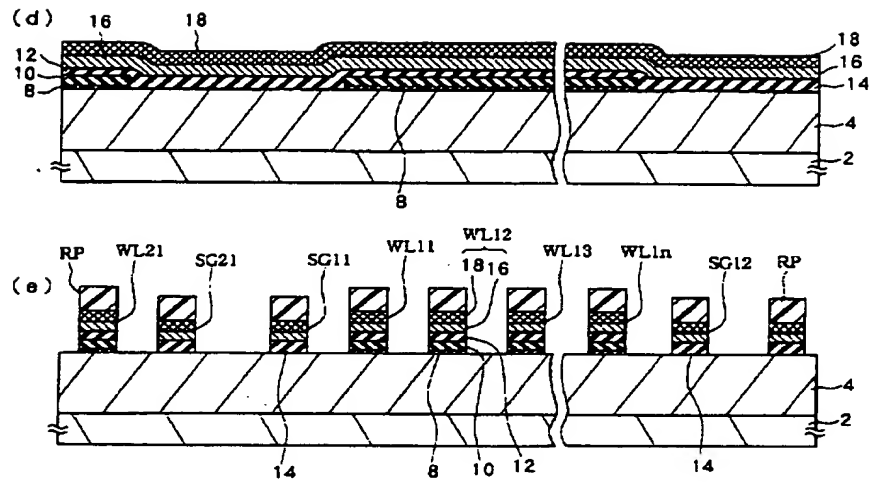
WL12 (第1の制御電極) 22第2の制御電極



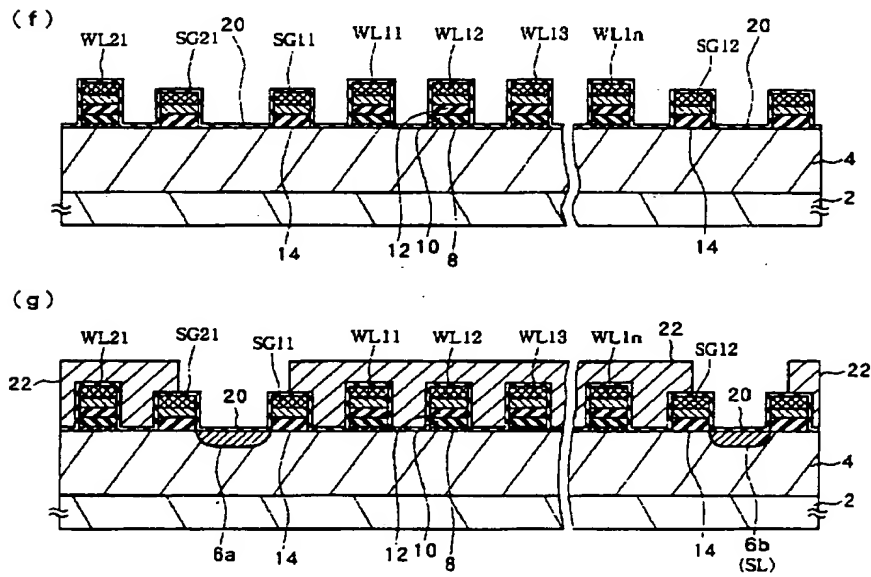
(a)



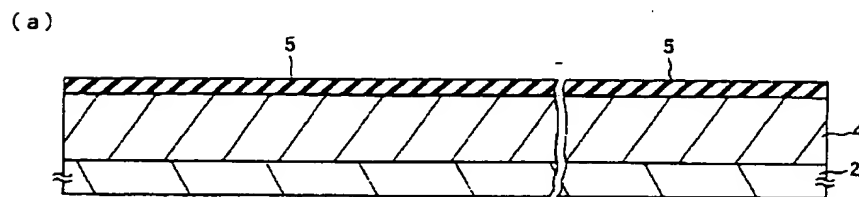
【図 5】



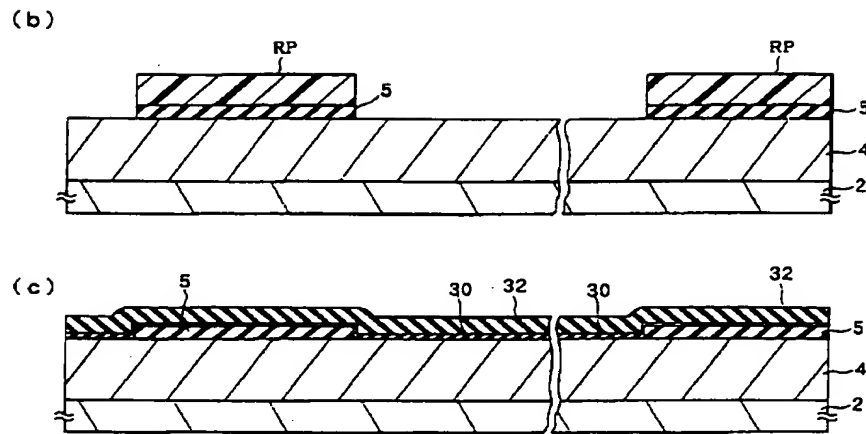
【図 6】



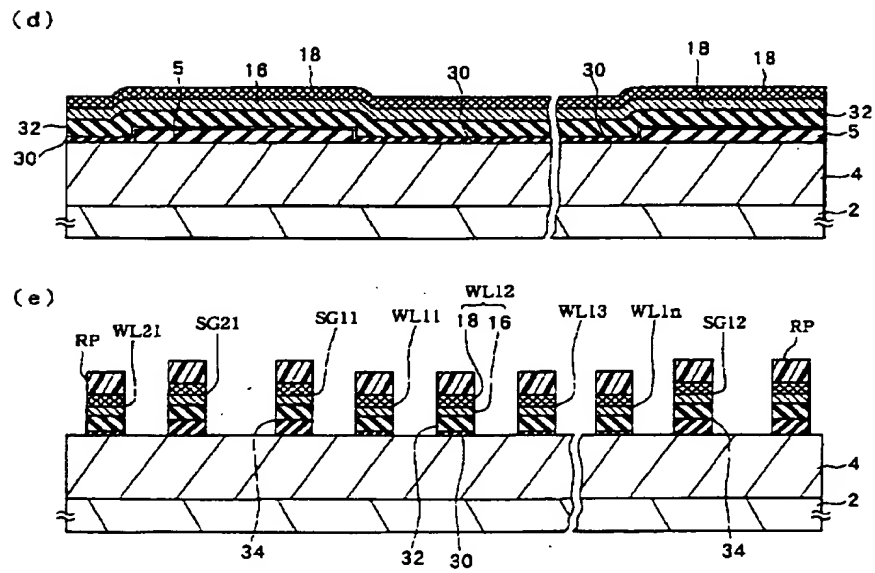
【図 8】



【図 9】

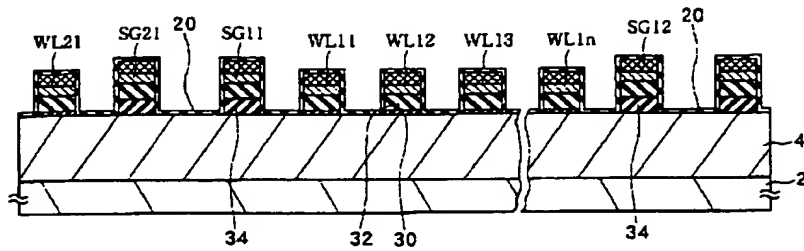


【図 10】

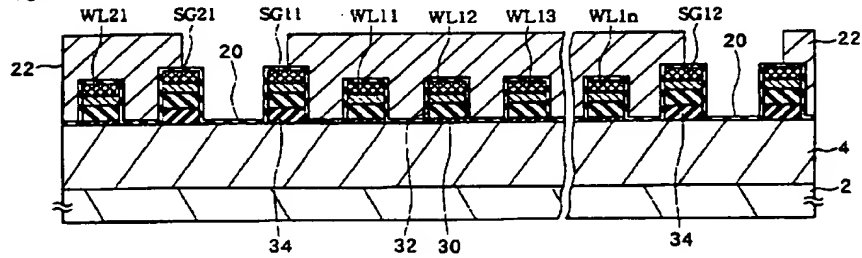


【図 1 1】

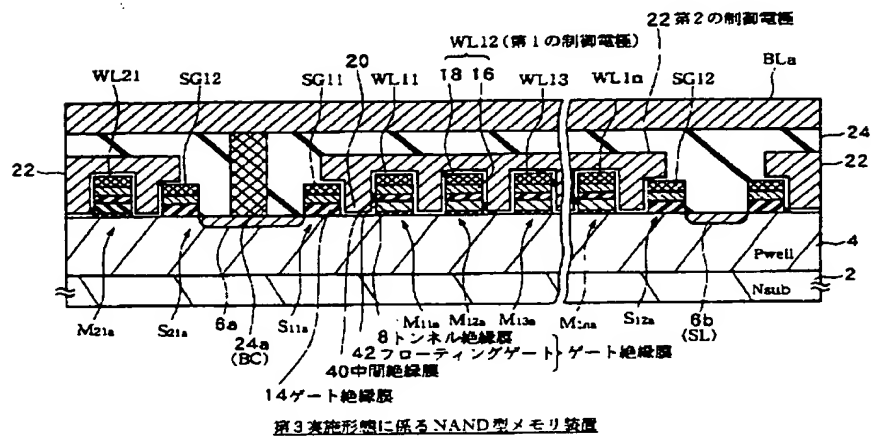
(f)



(g)



【図 1 2】



[illegible]